

CC derived oligopeptides of the invention

XX SQ Sequence 10 AA;

Query Match 100.0%; Score 62; DB 4; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.023;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGHPRPGR 10

Db 1 GGHPRPGR 10

RESULT 2

AA48783 ID AAB48783 standard; peptide; 17 AA.

XX AC

XX AAB48783;

XX 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 99-115), SEQ ID NO:13.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;

XX chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

XX oral bacterium; caries prevention.

XX OS Homo sapiens.

XX WO200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-S000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO/) STROEMBERG N.

XX (JOHA/) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of

XX proline-rich proteins, useful for preventing dental caries.

XX Claim 2; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-

XX AAB48783) which contain at least two arginine residues and which protect

XX against dental caries. PRPs (proline-rich proteins) are salivary proteins

XX encoded by six clustered genes on chromosome 12p13.2 and are potential

XX determinants of a person's susceptibility to dental caries. PRPs are

XX degraded by Actinomyces and Streptococcus species to small peptide

XX fragments. These are metabolised by oral bacteria for nutritional

XX purposes, with certain bacterial species generating ammonia via the

XX catabolism of arginine. The peptides of the invention, being arginine-

XX rich, can also be converted to ammonia by these bacteria. The ammonia

XX thus formed raises the pH at the dental surface, thereby protecting the

XX teeth against caries. Sequences AAB48771-48783 represent the PRP-1-

XX derived oligopeptides of the invention

XX SQ Sequence 17 AA;

Query Match 100.0%; Score 62; DB 4; Length 17;

Best Local Similarity 100.0%; Pred. No. 0.037;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGHPRPGR 10

Db 1 GGHPRPGR 10

RESULT 3

ABB3848 ID

ABB3848 standard; peptide; 132 AA.

XX AC

ABB3848;

XX 04-FEB-2002 (first entry)

XX Peptide #6354 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX OS Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000669.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human foetal liver.

XX Claim 27; SEQ ID NO 31483; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring

XX human gene expression in a sample derived from human foetal liver. The

XX single exon nucleic acid probes may be used for predicting, measuring and

XX displaying gene expression in samples derived from human foetal liver. The

XX present sequence is a peptide encoded by a single exon nucleic acid probe

XX of the invention. Note: The sequence data for this patent did not form

XX part of the printed specification, but was obtained in electronic format

XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 62; DB 4; Length 132;

Best Local Similarity 100.0%; Pred. No. 0.23;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GGHPRPGR 10

Db 81 GGHPRPGR 90

RESULT 4

AA32323 ID

AA32323 standard; protein; 132 AA.

XX AC

AA32323;

XX 17-OCT-2001 (first entry)

XX Peptide #6360 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder.

XX KW

XX

```

OS Homo sapiens.
XX WO200157272-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000663.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-48897/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX Claim 27; SEQ ID NO 32592; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP.
XX see AI1315-AA157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX human placenta.
XX Sequence 132 AA;
XX
XX Query Match 100.0%; Score 62; DB 4; Length 132;
XX Best Local Similarity 100.0%; Pred. No. 0.23;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GGHRPPRGR 10
XX 81 GGHRPPRGR 90
XX
XX RESULT 5
XX AM72058
XX ID AM72058 standard; protein; 132 AA.
XX AC AM72058;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 32364.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX WO200157276-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000668.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.

```

```

XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX Example 4; SEQ ID NO 32364; 658pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention
XX Sequence 132 AA;
XX
XX Query Match 100.0%; Score 62; DB 4; Length 132;
XX Best Local Similarity 100.0%; Pred. No. 0.23;
XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 GGHRPPRGR 10
XX 81 GGHRPPRGR 90
XX
XX RESULT 6
XX AAM59494
XX ID AAM59494 standard; protein; 132 AA.
XX AC AAM59494;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 31599.
XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;
XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX OS Homo sapiens.
XX WO200157275-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000667.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX Example 4; SEQ ID NO 31599; 650pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,

```

CC which may enable the diagnosis and improved treatment of nervous system  
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
CC epilepsy and cancers. The present sequence is a protein encoded by one of  
CC the probes of the invention

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 62; DB 4; Length 132;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRPPRGR 10  
| | | | | | | | | |  
DB 81 GGHPRPPRGR 90

RESULT 7  
ABG53742  
ID ABG53742 standard; peptide; 132 AA.

XX AC ABG53742;

XX DT 25-FEB-2003 (first entry)

XX DE Human liver peptide, SEQ ID No 32390.

XX XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
KW hypercholesterolaemia; coronary heart disease.

XX OS Homo sapiens.

XX PN WO200157273-A2.

XX XX 09-AUG-2001.

XX PF 30-JAN-2001; 2001WO-US000664.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2001-488898/53.

XX PT Human genome-derived single exon nucleic acid probes useful for analyzing

XX PS gene expression in human adult liver.

XX PS Claim 27; SEQ ID NO 32390; 658pp; English.

XX CC The invention relates to a single exon nucleic acid probe (SENP) (I) for  
CC measuring human gene expression in a sample derived from human adult  
CC liver, comprising one of 13109 defined nucleotide sequences given in the  
CC specification (or complements/ fragments). The probe hybridises at high  
CC stringency to a nucleic acid molecule expressed in the human adult liver.  
CC (I) may be used for predicting, measuring and displaying gene expression  
CC in samples derived from human adult liver. The genes identified may be  
CC involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
CC associated with coronary heart disease. ABG47348-ABG59930 represent human  
CC liver single exon encoded peptides of the invention. Note: The sequence  
CC information for this patent does not appear in the printed specification  
CC but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 62; DB 4; Length 132;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRPPRGR 10  
| | | | | | | | | |  
DB 81 GGHPRPPRGR 90

RESULT 8  
ABG41873  
ID ABG41873 standard; peptide; 132 AA.

XX AC ABG41873;

XX DT 19-AUG-2002 (first entry)

XX DE Human peptide encoded by genome-derived single exon probe SEQ ID 31538.

XX XX Human; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;  
KW pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
KW hyaline membrane disease.

XX OS Homo sapiens.

XX PN WO200186003-A2.

XX XX 15-NOV-2001.

XX PF 30-JAN-2001; 2001WO-US000665.

XX PR 04-FEB-2000; 2000US-0180312P.

XX PR 26-MAY-2000; 2000US-0207456P.

XX PR 30-JUN-2000; 2000US-00608408.

XX PR 03-AUG-2000; 2000US-00632366.

XX PR 21-SEP-2000; 2000US-0234687P.

XX PR 27-SEP-2000; 2000US-0236359P.

XX PR 04-OCT-2000; 2000GB-00024263.

XX PA (MOLE-) MOLECULAR DYNAMICS INC.

XX PI Penn SG, Hanzel DK, Chen W, Rank DR;

XX DR WPI; 2002-114183/15.

XX PT Spatially-addressable set of single exon nucleic acid probes, used to

XX PS measure gene expression in human lung samples.

XX PS Claim 27; SEQ ID NO 31538; 634pp; English.

XX CC The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of probes  
CC; the novel set of probes which hybridise at high stringency to a nucleic  
CC acid expressed in the human lung; measuring gene expression in a sample  
CC derived from human lung, comprising (a) contacting the array with a  
CC collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of the  
CC array; identifying exons in a eukaryotic genome, comprising (a)  
CC algorithmically predicting at least one exon from genomic sequences of  
CC the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,

comprising (a) identifying exons from genomic sequence by the method above and (b) measuring the expression of each of the exons in several tissues and/or cell types using hybridisation to a single exon microarrays having a probe with the exon, where a common pattern of expression of the exons in the tissues and/or cell types indicates that the exons should be assigned to a single gene; a peptide comprising one of 12011 sequences, mentioned in the specification, or encoded by the probes/open reading frames (ORF). The probes are used for gene expression analysis, and for identifying exons in a gene, particularly using human lung derived mRNA and for the study of lung diseases such as asthma, lung cancer, chronic obstructive pulmonary disease (COPD), interstitial lung disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis, tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Herxman-Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis, Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary dyskinesia, pulmonary hypertension and hyaline membrane disease. The present sequence is a peptide/protein encoded by a single exon probe of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WFO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 132 AA;

Query Match 100.0%; Score 62; DB 5; Length 132;  
Best Local Similarity 100.0%; Pred. No. 0.23;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHRPPRGR 10  
DB 81 GGHRPPRGR 90

RESULT 9  
ABR57423  
ID ABR57423 standard; protein; 149 AA.

AC ABR57423;

DT 15-SEP-2003 (first entry)

DE Human NOV7 protein SEQ ID NO:24.

XX Human; NOVX; cytostatic; cardiant; antiinflammatory; immunosuppressive;  
KW antiallergic; haemostatic; anti-HIV; antidiabetic; antiarteriosclerotic;  
KW anorectic; antiasthmatic; nephrotropic; antiarthritic; hepatotropic;  
KW neuroprotective; nontropic; antibacterial; virucide; antiparasitic;  
KW relaxant; anticonvulsant; hypotensive; vasotropic; antiparkinsonian;  
KW vulnary; angiogenic; antiangiogenic; gene therapy; vaccine; cancer;  
KW cardiomyopathy; atherosclerosis; hypertension; diabetes; inflammation;  
KW autoimmune disorder; allergy; blood disorder; AIDS; obesity; asthma;  
KW acquired immunodeficiency syndrome; nephropathy; cirrhosis; arthritis;  
KW Alzheimer's disease; Parkinson's disease; Goitre; infection; stroke;  
KW muscular dystrophy; epilepsy; wasting disorder; chromosome 12.

XX Homo sapiens.

XX WO200294870-A2.

XX 28-NOV-2002.

XX 02-NOV-2001; 2001WO-US051580.

XX 02-NOV-2000; 2000US-0245291P.

XX 02-NOV-2000; 2000US-0245317P.

XX 07-NOV-2000; 2000US-0246562P.

XX 08-NOV-2000; 2000US-0246871P.

XX 26-JAN-2001; 2001US-0264389P.

XX 26-JAN-2001; 2001US-0264423P.

XX 29-JAN-2001; 2001US-0264799P.

XX (CURA-) CURAGEN CORP.

PI Grosse WM, Macdougall JR, Smithson G, Millet I, Stone DJ;  
PI Gunter E, Ellerman K, Alsobrook JP, Lepley DM, Burgess CE;  
PI Sytek KA, Edinger SR, Gangolli EA, Gorman L, Taupier RJ, Li L;  
PI Guo X, Fernandes ER, Vernet CAM, Tchernev VT, Casman SJ, Shenoy S;  
PI Mishra V, Furtak K, Baumgartner JC, Colman SD;  
XX WPI: 2003-140359/13.  
DR N-PSDB; ACF03558.  
XX New NOVX polypeptide useful for preventing or treating NOVX-associated  
PT disorders, e.g. cancer, cardiomyopathy, atherosclerosis or diabetes, and  
PT in chromosome mapping, tissue typing or pharmacogenomics.

Claim 1: Page 69; 346pp; English.

XX ACF03547 to ACF03570 encode the human NOVX proteins (I) given in ABR57412  
CC to ABR57435. (I) have cytostatic, cardiant, antiinflammatory, nontropic,  
CC immunosuppressive, antiallergic, haemostatic, anti-HIV, antidiabetic,  
CC antiarteriosclerotic, anorectic, antiasthmatic, nephrotropic, virucide,  
CC antiarthritic, hepatotropic, neuroprotective, antibacterial, relaxant,  
CC antiparasitic, anticonvulsant, hypotensive, vasotropic, antiparkinsonian,  
CC vulnary, angiogenic and antiangiogenic activities, and can be used in  
CC gene therapy and vaccines. The NOVX polypeptides and their antibodies can  
CC be used to determine the presence or absence of (I) in a sample. The NOVX  
CC polypeptides, polynucleotides encoding them, and antibodies against them,  
CC are useful in manufacturing a medicament for treating or preventing a  
CC syndrome associated with a NOVX-associated disorder such as hypertension,  
CC cardiomyopathy, atherosclerosis, cancer, diabetes, asthma, inflammation,  
CC autoimmune disorders, allergies, blood disorders, obesity, acquired  
CC immunodeficiency syndrome (AIDS), immunoglobulin (Ig)A nephropathy,  
CC cirrhosis, arthritis, Alzheimer's disease, Parkinson's disease, goitre,  
CC infections (e.g. bacterial, viral, parasitic), stroke, muscular  
CC dystrophy, epilepsy, and other wasting disorders associated with chronic  
CC diseases. ACF03571 to ACF03644 represent PCR primers and probes for NOVX  
CC sequence, which are used in an example from the present invention

XX Sequence 149 AA;

Query Match 100.0%; Score 62; DB 6; Length 149;

Best Local Similarity 100.0%; Pred. No. 0.26;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHRPPRGR 10  
DB 98 GGHRPPRGR 107

RESULT 10

ABR56769

ID ABR56769 standard; protein; 154 AA.

XX ABR56769;

DT 30-JUL-2003 (first entry)

XX Human secreted protein SECP-44 SEQ ID NO:44.

XX Human; secreted protein; SECP; cytostatic; antiarteriosclerotic;  
KW anticonvulsant; nontropic; neuroprotective; cerebroprotective; anti-HIV;  
KW antiallergic; antiinflammatory; thyromimetic; gene therapy; cancer;  
KW cell proliferative disorder; atherosclerosis; neurological disorder;  
KW epilepsy; Huntington's disease; stroke; immune disorder; allergy; AIDS;  
KW inflammatory disorder; developmental disorder; hypothyroidism;  
KW Cushing's syndrome; infection.

XX Homo sapiens.

XX WO2003016506-A2.

XX 27-FEB-2003.

XX 15-AUG-2002; 2002WO-US027143.



PR 17-AUG-2001; 2001US-0313249P.  
PR 24-AUG-2001; 2001US-0314752P.  
PR 07-SEP-2001; 2001US-0317818P.  
PR 21-SEP-2001; 2001US-0317824P.  
PR 24-SEP-2001; 2001US-0324040P.  
PR 28-SEP-2001; 2001US-0324586P.  
PR 28-NOV-2001; 2001US-0343980P.  
PR 13-FEB-2002; 2002US-0357002P.  
PR 06-MAR-2002; 2002US-0362439P.  
PR 19-MAR-2002; 2002US-0366041P.  
PR 30-APR-2002; 2002US-0376988P.  
XX (INCY-) INCYTE GENOMICS INC.  
XX Tang TY, Warren BA, Gierzen KJ, Lal PG, Yue H, Honchell CD;  
PI Lehr-Mason PM, Burford N, Xu Y, Baughn MR, Duggan BM, Tran UK;  
PI Lee EA, Forsythe IU, Richardson TW, Lee S, Thangavelu K, Yue H;  
PI Emerling BM, Walia NK, Azimzai Y, Sanjanwala B, Hafalia AaA;  
PI Borowsky ML, Nguyen DB, Ison DB, Astromoff A, Ding L, Lee SY;  
PI Becha SD, Ramkumar J, Gandhi AS, Jin P, Fu GK, Swarnakar A;  
XX WPI: 2003-278569/27.  
DR N-PSDB; ACC79069.  
XX  
XX New human secreted proteins (SECP), useful for diagnosing, treating and  
PT preventing diseases or conditions associated with the aberrant SECP  
PT expression e.g. cancer, AIDS, atherosclerosis, Huntington's disease,  
PT stroke, infections.  
XX  
XX Claim 1; Page 222; 286pp; English.  
PS  
XX ACC79026 to ACC79105 encode the human secreted proteins (I) given in  
CC ABR56726 to ABR56805, designated SECP-1 to SECP-80. SECP sequences can  
CC have cytotoxic, antiarteriosclerotic, anticonvulsant, antiinflammatory,  
CC neurocytic, neuroprotective, cerebroprotective, anti-HIV, antiallergic and  
CC thrombolytic activities, and can be used in gene therapy. The SECP  
CC proteins and polynucleotides can be used in diagnosing, treating and  
CC preventing diseases or conditions associated with the decreased  
CC expression or overexpression of SECP, such as cell proliferative (e.g.  
CC cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's  
CC disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and  
CC developmental (e.g. hypothyroidism, Cushing's syndrome) disorders, or  
CC infections. They are also useful in assessing the effects of exogenous  
CC compounds on the expression of nucleic acid and amino acid sequences of  
CC SECP. The SECP or its fragments are useful in screening compounds for  
CC effectiveness as agonist or antagonist of the polypeptides, or in  
CC altering the expression of the target polynucleotide and compounds that  
CC specifically bind to or modulate the activity of the polypeptide  
XX  
XX Sequence 154 AA;  
Query Match 100.0%; Score 62; DB 6; Length 154;  
Best Local Similarity 100.0%; Pred. No. 0.27;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGHPRPPRGR 10  
Db 103 GGHPRPPRGR 112  
RESULT 11  
ADA83798  
ID ADA83798 standard; protein; 166 AA.  
XX  
XX ADA83798;  
AC  
XX  
XX 20-NOV-2003 (first entry)  
DT  
XX  
XX Human PRF2 protein.  
DE  
XX human; marker; expressed sequence tag; EST; arabidopsis; tumour;  
XX stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;  
KW

KW vaccine.  
XX  
XX Homo sapiens.  
XX WO2002103028-A2.  
XX  
XX 27-DEC-2002.  
XX  
XX 30-MAY-2002; 2002WO-1E004189.  
XX  
XX 30-MAY-2001; 2001US-0293999P.  
PR 22-OCT-2001; 2001US-0330457P.  
PR 19-FEB-2002; 2002US-0357144P.  
XX  
XX (BIOM-) BIOMEDICAL CENT.  
XX Baranova AV, Yankovsky NK, Kozlov AP, Lobashev AV, Krukovskaya LL;  
XX WPI: 2003-175241/17.  
DR N-PSDB; ADA83797.  
XX  
XX Determining if a nucleic acid is a marker for a phenotype/cell type of  
PT interest, by global comparison of expressed sequence tags known to be  
PT expressed in the phenotype/cell type with all ESTs expressed in normal  
PT tissue.  
XX  
XX Claim 29; Page 191-192; 516pp; English.  
PS  
XX The invention relates to a novel method for determining if a nucleic acid  
CC is a marker for a predetermined phenotype/cell type of interest from a  
CC biological species. The method comprises performing a global comparison  
CC of a group of expressed sequence tags (ESTs) known to be expressed in the  
CC phenotype/cell type of interest with all ESTs expressed in normal tissue  
CC in order to identify ESTs that are preferentially expressed in the  
CC phenotype/cell of interest. A method of the invention is useful for  
CC determining whether a nucleic acid is a marker for a predetermined  
CC phenotype or cell type of interest from a biological species, preferably  
CC Arabidopsis or human. The cell type of interest is an abnormal cell such  
CC as a tumour cell, and the predetermined phenotype is a stress-induced  
CC phenotype such as hyperosmotic stress or high salt conditions. A method  
CC of the invention is also useful for determining the progression of colon  
CC cancer in a human, for detecting a tumour cell, and for regulating or  
CC preventing the growth of a tumour cell. An antibody of the invention is  
CC useful for detecting the absence or presence of peptides encoded by  
CC tumour-associated markers. A polypeptide of the invention is useful as an  
CC immunogen for vaccinating an animal. The present sequence represents a  
CC tumour-associated antigen of the invention.  
XX  
XX Sequence 166 AA;  
Query Match 100.0%; Score 62; DB 6; Length 166;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GGHPRPPRGR 10  
Db 115 GGHPRPPRGR 124  
RESULT 12  
ADC98216  
ID ADC98216 standard; protein; 166 AA.  
XX  
XX ADC98216;  
AC  
XX  
XX 01-JAN-2004 (first entry)  
DT  
XX  
XX Human salivary acidic proline-rich phosphoprotein (PRP).  
DE  
XX  
XX Human; salivary acidic proline-rich phosphoprotein; PRP; autoantigen;  
KW atopic dermatitis-inducing protein; salivary gland; IGE autoantibody;  
KW immunoglobulin E; mast cell activation; basophil activation; diagnosis;  
KW risk assessment; sensitisation remedy; dermatological; antiallergic;  
KW

KW antiinflammatory.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003084991-A1.  
 XX  
 PD 16-OCT-2003.  
 XX  
 XX 04-APR-2003; 2003WO-JP004325.  
 XX  
 PF 08-APR-2002; 2002JP-00105425.  
 XX  
 XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
 PA  
 PI Hide M, Yamamoto S, Tanaka T, Koro O;  
 DR WPI; 2003-833567/77.  
 DR N-PSDB; ADC98215.  
 XX  
 PT Atopic dermatitis-inducing proteins, applicable in diagnosis of including  
 PT risk of onset, and in developing sensitization remedies for the disease.  
 XX  
 PS Claim 4; SEQ ID NO 2; 43pp; Japanese.  
 XX  
 CC The invention relates to the human atopic dermatitis-inducing proteins,  
 CC salivary acidic proline-rich phosphoprotein (PRP; ADC98216) and prolactin  
 CC -inducible protein (PIP; ADC98218), and their post-translationally  
 CC modified forms. These proteins are secreted by salivary or sweat glands  
 CC and bind to IgE autoantibodies, thereby activating mast cells and  
 CC basophils. The invention also relates to antigenic peptide fragments of  
 CC PRP or PIP; antibodies which bind to PRP or PIP; methods for diagnosing  
 CC atopic dermatitis or for determining whether an individual is at risk of  
 CC developing atopic dermatitis by determining the presence of PRP- or PIP-  
 CC specific antibodies or immune complexes, or by quantifying histamine  
 CC release; and sensitization remedies for atopic dermatitis containing PRP  
 CC and/or PIP or their peptides as the active ingredient. PRP, PIP and their  
 CC antibodies are useful in diagnosing atopic dermatitis, or for determining  
 CC whether an individual is at risk of developing atopic dermatitis. They  
 CC are also useful in developing sensitization remedies for the treatment of  
 CC atopic dermatitis. The present sequence represents the specifically  
 CC claimed human salivary acidic proline-rich phosphoprotein (PRP).  
 XX  
 SQ Sequence 166 AA;  
 Query Match 100.0%; Score 62; DB 7; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 0.28;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGHPRPPRCR 10  
 DB |||||||||  
 115 GGHPRPPRCR 124  
 RESULT 13  
 AAB48778  
 ID AAB48778 standard; peptide; 9 AA.  
 XX  
 AC AAB48778;  
 XX  
 XX 09-MAR-2001 (first entry)  
 DT  
 XX Human saliva PRP-1 fragment (residues 100-108), SEQ ID NO:8.  
 DE  
 XX Human; PRP-1; proline-rich protein; saliva; dental caries;  
 XX chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200069890-A1.  
 PN  
 XX 23-NOV-2000.  
 PD  
 XX

PF 11-MAY-2000; 2000WO-SE000930.  
 XX  
 PR 17-MAY-1999; 99SE-00001773.  
 XX  
 PA (STRO/) STROEMBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX  
 PI Stroemberg N, Johansson I;  
 XX  
 DR WPI; 2001-031923/04.  
 XX  
 PT New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 XX  
 PS Claim 4; Page 24; 36pp; English.  
 XX  
 CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention  
 XX  
 SQ Sequence 9 AA;  
 Query Match 90.3%; Score 56; DB 4; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GHPRRPRGR 10  
 DB |||||||||  
 1 GHPRRPRGR 9  
 RESULT 14  
 ABE28896  
 ID ABE28896 standard; protein; 843 AA.  
 XX  
 AC ABE28896;  
 XX  
 DT 18-FEB-2002 (first entry)  
 DT  
 XX Novel human diagnostic protein #28887.  
 DE  
 XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.  
 KW  
 XX Homo sapiens.  
 OS  
 XX WO200175067-A2.  
 PN  
 XX 11-OCT-2001.  
 PD  
 XX 30-MAR-2001; 2001WO-US008631.  
 PF  
 XX 31-MAR-2000; 2000US-00540217.  
 PR  
 XX 23-AUG-2000; 2000US-00849167.  
 PR  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX Drmanac RT, Liu C, Tang YT;  
 PI  
 XX WPI; 2001-639362/73.  
 DR  
 XX N-PSDB; AAS93083.  
 DR  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT

PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

PS Claim 20; SEQ ID NO 59255; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 843 AA;

Query Match 82.3%; Score 51; DB 4; Length 843;  
 Best Local Similarity 100.0%; Pred. No. 42;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGHPRP RP 8  
 Db 20 GGHPRP RP 27

RESULT 15  
 AAB48779  
 ID AAB48779 standard; peptide; 8 AA.  
 XX  
 AC AAB48779;  
 XX  
 DT 09-MAR-2001 (first entry)  
 XX  
 DE Human saliva PRP-1 fragment (residues 101-108), SEQ ID NO:9.  
 XX  
 KW Human; PRP-1; proline-rich protein; saliva; dental caries;  
 KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200069890-A1.  
 XX  
 PD 23-NOV-2000.  
 XX  
 PF 11-MAY-2000; 2000WO-SE000930.  
 XX  
 PR 17-MAY-1999; 99SE-00001773.  
 XX  
 PA (STRO/) STROEMBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX  
 PI Stroemberg N, Johansson I;  
 XX  
 DR WPI; 2001-031923/04.  
 XX  
 PT New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 XX

PS Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention

XX SQ Sequence 8 AA;

Query Match 80.6%; Score 50; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HRP RP RP 10  
 Db 1 HRP RP RP 8

Search completed: April 6, 2004, 16:06:39  
 Job time : 56.9813 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 16:14:50 ; Search time 39.2523 Seconds  
(without alignments)  
66.909 Million cell updates/sec

Title: US-10-009-709-7  
Perfect score: 62  
Sequence: 1 GGHPRPPRGR 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 26263353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*

```

1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	62	100.0	132	9	US-09-864-761-43644
2	62	100.0	166	14	Sequence 43644, A
3	48	77.4	19662	15	Sequence 80, Appl
4	46	74.2	326	9	US-10-084-846A-6
5	46	74.2	326	9	Sequence 6, Appl
6	46	74.2	19652	15	Sequence 10259, A
7	45	72.6	238	12	Sequence 14087, A
8	45	72.6	277	12	Sequence 7, Appl
9	44	71.0	72	12	Sequence 69167, A
10	43.5	70.2	707	10	Sequence 43478, A
11	43	69.4	123	12	Sequence 278, App
12	43	69.4	185	12	Sequence 274737, A
13	43	69.4	191	15	Sequence 250895, A
14	42	67.7	40	9	US-10-374-780A-1418
15	42	67.7	229	12	Sequence 45389, A
					Sequence 149900, A

16	42	67.7	230	12	US-10-424-599-233437
17	42	67.7	263	12	Sequence 233437, A
18	42	67.7	263	12	Sequence 50588, A
19	42	67.7	263	12	Sequence 60148, A
20	42	67.7	265	12	Sequence 60768, A
21	42	67.7	383	12	Sequence 48471, A
22	42	67.7	383	12	Sequence 64314, A
23	42	67.7	81	12	Sequence 73016, A
24	41	66.1	94	12	Sequence 237492, A
25	41	66.1	104	12	Sequence 285441, A
26	41	66.1	112	12	Sequence 178384, A
27	41	66.1	133	12	Sequence 182615, A
28	41	66.1	226	12	Sequence 278333, A
29	41	66.1	433	12	Sequence 64280, A
30	41	66.1	466	9	Sequence 336, App
31	41	66.1	466	14	Sequence 1355, App
32	41	66.1	550	14	Sequence 6280, App
33	41	66.1	595	9	Sequence 3280, App
34	41	66.1	687	10	Sequence 67, Appl
35	41	66.1	687	12	Sequence 18, Appl
36	41	66.1	687	14	Sequence 584, App
37	41	66.1	687	15	Sequence 76, Appl
38	41	66.1	747	15	Sequence 174, App
39	41	66.1	924	12	Sequence 2445, App
40	41	66.1	1471	8	Sequence 63880, A
41	41	66.1	1474	14	Sequence 1, Appl
42	41	66.1	1474	15	Sequence 522, App
43	41	66.1	1515	14	Sequence 914, App
44	41	66.1	2796	9	Sequence 8, Appl
45	41	66.1	2796	10	Sequence 114, App
					Sequence 114, App

## ALIGNMENTS

## RESULT 1

```

US-09-864-761-43644
; Sequence 43644, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David X.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30

```

```
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43644
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006519.17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77
; OTHER INFORMATION: EST HUMAN HIT: BF088785.1, EVALUE 1.00e-01
; OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02
US-09-864-761-43644
```

```
Query Match 100.0%; Score 62; DB 9; Length 132;
Best Local Similarity 100.0%; Pred. No. 0.5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 GGHPRPPRGR 10
Db 81 GGHPRPPRGR 90
```

## RESULT 2

```
US-10-157-031-80
; Sequence 80, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krukovskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-157-031-80
```

```
Query Match 100.0%; Score 62; DB 14; Length 166;
Best Local Similarity 100.0%; Pred. No. 0.61;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 GGHPRPPRGR 10
Db 115 GGHPRPPRGR 124
```

## RESULT 3

```
US-10-084-846A-6
; Sequence 6, Application US/10084846A
```

```
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MOHLENWEG, AGNES
; APPLICANT: TREFFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 6
; LENGTH: 19662
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 2.
; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.
US-10-084-846A-6
```

```
Query Match 77.4%; Score 48; DB 15; Length 19662;
Best Local Similarity 80.0%; Pred. No. 2.8e+03;
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy 1 GGHPRPPRGR 10
Db 13061 GGHPRPPRGR 13070
```

## RESULT 4

```
US-09-815-242-10259
; Sequence 10259, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlser, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10259
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Escherichia coli
US-09-815-242-10259
```

Query Match 74.2%; Score 46; DB 9; Length 326;  
Best Local Similarity 77.8%; Pred. No. 1.4e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGRPRPPRG 9  
|||  
Db 254 GGRPRPPKG 262

## RESULT 5

US-09-815-242-14087  
; Sequence 14087, Application US/09815242  
; Patent No. US20020061569A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Orisen, Kari L.  
; APPLICANT: Zyskind, Judith W.  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John D.  
; APPLICANT: Carr, Grant J.  
; APPLICANT: Yamamoto, Robert T.  
; APPLICANT: Xu, H. Howard

; TITLE OF INVENTION: Identification of Essential Genes in  
; Prokaryotes

; FILE REFERENCE: ELITRA.011A

; CURRENT APPLICATION NUMBER: US/09/815,242

; CURRENT FILING DATE: 2001-03-21

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/131,078

; PRIOR FILING DATE: 2000-03-21

; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26

; PRIOR APPLICATION NUMBER: 60/242,578

; PRIOR FILING DATE: 2000-10-23

; PRIOR APPLICATION NUMBER: 60/253,625

; PRIOR FILING DATE: 2000-11-27

; PRIOR APPLICATION NUMBER: 60/257,931

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16

; NUMBER OF SEQ ID NOS: 14110

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 14087

; LENGTH: 326

; TYPE: PRT

; ORGANISM: Salmonella typhi

; FEATURE:

; NAME/KEY: VARIANT

; LOCATION: (1)...(326)

; OTHER INFORMATION: Xaa = Any Amino Acid

US-09-815-242-14087

Query Match 74.2%; Score 46; DB 9; Length 326;  
Best Local Similarity 77.8%; Pred. No. 1.4e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1' GGRPRPPRG 9  
|||  
Db 254 GGRPRPPKG 262

## RESULT 6

US-10-084-846A-7

; Sequence 7, Application US/10084846A

; Publication No. US2004006026A1

; GENERAL INFORMATION:

; APPLICANT: WEITNAUER, GABRIELE

; APPLICANT: MOHLENWEG, AGNES

; APPLICANT: TREFZER, AXEL

; APPLICANT: BECHTOLD, ANDREAS

; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES

; FILE REFERENCE: 1974-005  
; CURRENT APPLICATION NUMBER: US/10/084,846A  
; CURRENT FILING DATE: 2003-02-25  
; PRIOR APPLICATION NUMBER: PCT/EP01/09815  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: DE 101 09 166.4  
; PRIOR FILING DATE: 2001-02-25  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 7  
; LENGTH: 19652  
; TYPE: PRT  
; ORGANISM: Streptomyces viridochromogenes  
; FEATURE:

; OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 2.  
; OTHER INFORMATION: Start codon: gat, Start position: nucleotide 2.  
US-10-084-846A-7

Query Match 74.2%; Score 46; DB 15; Length 19652;  
Best Local Similarity 88.9%; Pred. No. 5.2e+03;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GHRPRPRGR 10  
|||  
Db 2258 GHRPRPRGR 2266

## RESULT 7

US-10-425-114-69167

; Sequence 69167, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E.

; APPLICANT: Tabaska, Jack E.

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 69167

; LENGTH: 238

; TYPE: PRT

; ORGANISM: Zea mays

; FEATURE:

; OTHER INFORMATION: Clone ID: UC-ZMFLMO17307C10\_FLI.pep

US-10-425-114-69167

Query Match 72.6%; Score 45; DB 12; Length 238;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HPRPPRG 9  
|||  
Db 155 HPRPPRG 161

## RESULT 8

US-10-425-114-43478

; Sequence 43478, Application US/10425114

; Publication No. US20040034888A1

; GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E.

; APPLICANT: Tabaska, Jack E.

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

;; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
;; FILE REFERENCE: 38-21(53313)B  
;; CURRENT APPLICATION NUMBER: US/10/425,114  
;; CURRENT FILING DATE: 2003-04-28  
;; NUMBER OF SEQ ID NOS: 73128  
;; SEQ ID NO 43478  
;; LENGTH: 277  
;; TYPE: PRT  
;; ORGANISM: Zea mays  
;; FEATURE:  
;; OTHER INFORMATION: Clone ID: 700423863\_FLI.pcp  
US-10-425-114-43478

Query Match 72.6%; Score 45; DB 12; Length 277;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02; Indels 0; Gaps 0;  
Matches 7; Conservative 0; Mismatches 0;

Qy 3 HRPFRPG 9  
Db 86 HRPFRPG 92

RESULT 9  
US-10-424-599-223703  
; Sequence 223703, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 223703  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_44033C.1.pcp  
US-10-424-599-223703

Query Match 71.0%; Score 44; DB 12; Length 72;  
Best Local Similarity 80.0%; Pred. No. 70;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGHPRPPRGR 10  
Db 24 GTRPRPPRGR 33

RESULT 10  
US-09-919-039-278  
; Sequence 278, Application US/09919039  
; Publication No. US20030108871A1  
; GENERAL INFORMATION:  
; APPLICANT: Kasey, Matthew R.  
; TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES  
; FILE REFERENCE: PA-0035 US  
; CURRENT APPLICATION NUMBER: US/09/919,039  
; CURRENT FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: 60/222,113  
; PRIOR FILING DATE: 2000-07-28  
; NUMBER OF SEQ ID NOS: 401  
; SOFTWARE: PERL Program  
; SEQ ID NO 278  
; LENGTH: 707  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:

;; NAME/KEY: misc feature  
;; OTHER INFORMATION: Incyte ID No. US20030108871A1 2457215CD1  
US-09-919-039-278

Query Match 70.2%; Score 43.5; DB 10; Length 707;  
Best Local Similarity 80.0%; Pred. No. 6e+02; Indels 1; Gaps 1;  
Matches 8; Conservative 1; Mismatches 0;

Qy 1 GGHPRPP-RG 9  
Db 228 GGHPRPPRGR 237

RESULT 11  
US-10-424-599-274737  
; Sequence 274737, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 274737  
; LENGTH: 123  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(123)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_90108C.1.pcp  
US-10-424-599-274737

Query Match 69.4%; Score 43; DB 12; Length 123;  
Best Local Similarity 85.7%; Pred. No. 1.5e+02; Indels 0; Gaps 0;  
Matches 6; Conservative 1; Mismatches 0;

Qy 1 GGHPRPP 7  
Db 15 GGHPRPP 21

RESULT 12  
US-10-424-599-250895  
; Sequence 250895, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 250895  
; LENGTH: 185  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(185)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:

OTHER INFORMATION: Clone ID: PAT\_MRT3847\_68588C.1.pcp  
US-10-424-599-250895

Query Match 69.4%; Score 43; DB 12; Length 185;  
Best Local Similarity 70.0%; Pred. No. 2.2e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGHPRPRGR 10  
DB 160 GGGPRPPACK 169

## RESULT 13

US-10-374-780A-1418  
; Sequence 1418, Application US/10374780A  
; Publication No. US20040019927A1  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, Bradley K  
; APPLICANT: Riechmann, Jose Luis  
; APPLICANT: Jiang, Cai-Zhong  
; APPLICANT: Heard, Jacqueline E  
; APPLICANT: Haake, Volker  
; APPLICANT: Creelman, Robert A  
; APPLICANT: Ratcliffe, Oliver  
; APPLICANT: Adam, Luc J  
; APPLICANT: Reuber, T. Lynne  
; APPLICANT: Keddle, James  
; APPLICANT: Broun, Pierre E  
; APPLICANT: Pilgrim, Marsha L  
; APPLICANT: Dubell III, Arnold T  
; APPLICANT: Pineda, Onaira  
; APPLICANT: Yu, Guo-Liang  
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS  
; FILE REFERENCE: MBI-0047 CIP  
; CURRENT APPLICATION NUMBER: US/10/374,780A  
; CURRENT FILING DATE: 2003-02-25  
; PRIOR APPLICATION NUMBER: 09/837,944  
; PRIOR FILING DATE: 2001-04-18  
; PRIOR APPLICATION NUMBER: 60/310,847  
; PRIOR FILING DATE: 2001-08-09  
; PRIOR APPLICATION NUMBER: 09/934,455  
; PRIOR FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: 60/336,049  
; PRIOR FILING DATE: 2001-11-19  
; PRIOR APPLICATION NUMBER: 60/338,692  
; PRIOR FILING DATE: 2001-12-11  
; PRIOR APPLICATION NUMBER: 10/171,468  
; PRIOR FILING DATE: 2002-06-14  
; PRIOR APPLICATION NUMBER: 10/225,066  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: 10/225,067  
; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: 10/225,068  
; PRIOR FILING DATE: 2002-08-09  
; NUMBER OF SEQ ID NOS: 2906  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1418  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Orthologous to G1073  
US-10-374-780A-1418

Query Match 69.4%; Score 43; DB 15; Length 191;  
Best Local Similarity 80.0%; Pred. No. 2.2e+02;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGHPRPRGR 10  
DB 50 GGGPRPRGR 59

## RESULT 14

US-09-864-761-45389  
; Sequence 45389, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-01-29  
; NUMBER OF SEQ ID NOS: 4917  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 45389  
; LENGTH: 40  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AP001427.1  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.77  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.75  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.55  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.71  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.61  
; OTHER INFORMATION: SWISSPROT HIT: O15016, EVALUE 3.70e+00  
; OTHER INFORMATION: EST\_HUMAN HIT: BF525926.1, EVALUE 2.40e+00  
US-09-864-761-45389

Query Match 67.7%; Score 42; DB 9; Length 40;  
Best Local Similarity 87.5%; Pred. No. 76;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GHHPRPRGR 9  
||| |||||



Db 31 GHRPPRG 38

RESULT 15  
US-10-424-599-149900  
; Sequence 149900, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 149900  
; LENGTH: 229  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(229)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_106381C.1.pap  
US-10-424-599-149900

Query Match 67.7%; Score 42; DB 12; Length 229;  
Best Local Similarity 70.0%; Pred. NO. 3.5e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GCHPPPPRG 10  
|||  
Db 93 GGTPPPRGK 102

Search completed: April 6, 2004, 17:05:59  
Job time : 44.2523 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 16:06:50 ; Search time 208.411 Seconds  
46.933 Million cell updates/sec

Title: US-10-009-709-7  
Perfect score: 62  
Sequence: 1 GHPRPGR 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 6019581 seqs, 976053577 residues

Total number of hits satisfying chosen parameters: 6019581

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Pending Parents AA Main:  
1: /cgn2\_6/ptodata/2/paa/US06\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/paa/US07\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/paa/US08\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/paa/US09\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/paa/US10\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/paa/US11\_COMB.pep.\*  
7: /cgn2\_6/ptodata/2/paa/US12\_COMB.pep.\*  
8: /cgn2\_6/ptodata/2/paa/US13\_COMB.pep.\*  
9: /cgn2\_6/ptodata/2/paa/US14\_COMB.pep.\*  
10: /cgn2\_6/ptodata/2/paa/US15\_COMB.pep.\*  
11: /cgn2\_6/ptodata/2/paa/US16\_COMB.pep.\*  
12: /cgn2\_6/ptodata/2/paa/US17\_COMB.pep.\*  
13: /cgn2\_6/ptodata/2/paa/US18\_COMB.pep.\*  
14: /cgn2\_6/ptodata/2/paa/US19\_COMB.pep.\*  
15: /cgn2\_6/ptodata/2/paa/US20\_COMB.pep.\*  
16: /cgn2\_6/ptodata/2/paa/US21\_COMB.pep.\*  
17: /cgn2\_6/ptodata/2/paa/US22\_COMB.pep.\*  
18: /cgn2\_6/ptodata/2/paa/US23\_COMB.pep.\*  
19: /cgn2\_6/ptodata/2/paa/US24\_COMB.pep.\*  
20: /cgn2\_6/ptodata/2/paa/US25\_COMB.pep.\*  
21: /cgn2\_6/ptodata/2/paa/US26\_COMB.pep.\*  
22: /cgn2\_6/ptodata/2/paa/US27\_COMB.pep.\*  
23: /cgn2\_6/ptodata/2/paa/US28\_COMB.pep.\*  
24: /cgn2\_6/ptodata/2/paa/US29\_COMB.pep.\*  
25: /cgn2\_6/ptodata/2/paa/US30\_COMB.pep.\*  
26: /cgn2\_6/ptodata/2/paa/US31\_COMB.pep.\*  
27: /cgn2\_6/ptodata/2/paa/US32\_COMB.pep.\*  
28: /cgn2\_6/ptodata/2/paa/US33\_COMB.pep.\*  
29: /cgn2\_6/ptodata/2/paa/US34\_COMB.pep.\*  
30: /cgn2\_6/ptodata/2/paa/US35\_COMB.pep.\*  
31: /cgn2\_6/ptodata/2/paa/US36\_COMB.pep.\*  
32: /cgn2\_6/ptodata/2/paa/US37\_COMB.pep.\*  
33: /cgn2\_6/ptodata/2/paa/US38\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match Length	DB ID	Description
------------	-------------	--------------	-------	-------------

1	62	100.0	10	26	US-10-009-709-7	Sequence 7, Appli
2	62	100.0	17	26	US-10-009-709-13	Sequence 13, Appl
3	62	100.0	128	23	US-09-834-366-17988	Sequence 17988, A
4	62	100.0	128	33	US-60-137-873-17988	Sequence 17988, A
5	62	100.0	131	23	US-09-834-366-13405	Sequence 13405, A
6	62	100.0	131	33	US-60-137-873-13405	Sequence 13405, A
7	62	100.0	132	1	PCT-US01-00663-32592	Sequence 32592, A
8	62	100.0	132	23	US-09-864-761-43644	Sequence 43644, A
9	62	100.0	132	27	US-10-182-993-31599	Sequence 31599, A
10	62	100.0	132	28	US-10-203-134-32364	Sequence 32364, A
11	62	100.0	132	28	US-10-203-135-31538	Sequence 31538, A
12	62	100.0	132	28	US-10-203-136-32390	Sequence 32390, A
13	62	100.0	132	28	US-10-203-137-32592	Sequence 32592, A
14	62	100.0	132	28	US-10-203-139-31483	Sequence 31483, A
15	62	100.0	149	26	US-10-002-945-24	Sequence 24, Appl
16	62	100.0	150	4	US-08-070-339A-5	Sequence 5, Appli
17	62	100.0	152	23	US-09-834-366-13417	Sequence 13417, A
18	62	100.0	152	33	US-60-137-873-13417	Sequence 13417, A
19	62	100.0	153	23	US-09-834-366-13409	Sequence 13409, A
20	62	100.0	153	33	US-60-137-873-13409	Sequence 13409, A
21	62	100.0	154	1	PCT-US02-27143-44	Sequence 44, Appl
22	62	100.0	166	21	US-09-724-676-62326	Sequence 62326, A
23	62	100.0	166	21	US-09-724-676A-62326	Sequence 62326, A
24	62	100.0	166	24	US-09-949-016-6709	Sequence 6709, Ap
25	62	100.0	166	26	US-10-002-945-85	Sequence 85, Appl
26	62	100.0	166	26	US-10-002-945-87	Sequence 87, Appl
27	62	100.0	166	26	US-10-002-945-88	Sequence 88, Appl
28	62	100.0	166	26	US-10-002-945-89	Sequence 89, Appl
29	62	100.0	166	27	US-10-157-031-80	Sequence 80, Appl
30	62	100.0	166	33	US-60-423-586-142	Sequence 142, App
31	62	100.0	166	33	US-60-427-194-142	Sequence 142, App
32	62	100.0	166	33	US-60-452-680-21393	Sequence 21393, A
33	62	100.0	166	33	US-60-490-890-1159	Sequence 1159, Ap
34	62	100.0	171	26	US-10-002-945-86	Sequence 86, Appl
35	62	100.0	178	24	US-09-949-016-8394	Sequence 8394, Ap
36	62	100.0	187	1	PCT-US03-10249-49	Sequence 49, Appl
37	62	100.0	202	22	US-09-760-477-406	Sequence 406, App
38	62	100.0	202	28	US-10-222-860-406	Sequence 406, App
39	56	90.3	9	26	US-10-009-709-5	Sequence 8, Appli
40	54	87.1	126	23	US-09-834-366-16601	Sequence 16601, A
41	54	87.1	126	33	US-60-137-873-16601	Sequence 16601, A
42	53	85.5	200	22	US-09-760-477-650	Sequence 650, App
43	53	85.5	200	28	US-10-222-860-650	Sequence 650, App
44	51	82.3	60	23	US-09-834-366-16373	Sequence 16373, A
45	51	82.3	60	33	US-60-137-873-16373	Sequence 16373, A

ALIGNMENTS

RESULT 1  
US-10-009-709-7  
; Sequence 7, Application US/10009709  
; GENERAL INFORMATION:  
; APPLICANT: Stromberg, Nicklas  
; APPLICANT: Ingegerd, Johansson  
; TITLE OF INVENTION: PREVENTION OF DENTAL CARRIES  
; FILE REFERENCE: P/2432-44  
; CURRENT APPLICATION NUMBER: US/10/009, 709  
; CURRENT FILING DATE: 2001-11-16  
; PRIOR APPLICATION NUMBER: PCT/SE00/00930  
; PRIOR FILING DATE: 1999-05-17  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; US-10-009-709-7

Query Match 100.0%; Score 62; DB 26; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.46; 0;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 GGHPRPPRGR 10
Db 1 GGHPRPPRGR 10

RESULT 2
US-10-009-709-13
; Sequence 13, Application US/10009709
; GENERAL INFORMATION:
; APPLICANT: Stromberg, Nicklas
; APPLICANT: Ingegard, Johansson
; TITLE OF INVENTION: PREVENTION OF DENTAL CARIES
; FILE REFERENCE: P/2432-44
; CURRENT APPLICATION NUMBER: US/10/009,709
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: PCT/SE00/00930
; PRIOR FILING DATE: 1999-05-17
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 17
; TYPE: PRT
; ORGANISM: Homo sapiens
; OTHER INFORMATION: Xaa = Ala, Asp, Gly, Val
US-10-009-709-13

Query Match 100.0%; Score 62; DB 26; Length 17;
Best Local Similarity 100.0%; Pred. No. 0.76; Length 128;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGHPRPPRGR 10
Db 1 GGHPRPPRGR 10

RESULT 3
US-09-834-366-17988
; Sequence 17988, Application US/09834366
; GENERAL INFORMATION:
; APPLICANT: Bejanin, Stephane
; APPLICANT: Tanaka, Hiroaki
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Jobert, Severin
; APPLICANT: Giordano, Jean-Yves
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: 81.US2.REG
; CURRENT APPLICATION NUMBER: US/09/834,366
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/197,873
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 52153
; SOFTWARE: Patent.pm
; SEQ ID NO 17988
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -16...-1
; NAME/KEY: UNSURE
; LOCATION: 100
; OTHER INFORMATION: Xaa = Ala, Asp, Gly, Val
US-09-834-366-17988

Query Match 100.0%; Score 62; DB 23; Length 128;
Best Local Similarity 100.0%; Pred. No. 4.9; Length 128;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGHPRPPRGR 10
Db 77 GGHPRPPRGR 86

RESULT 4
US-60-197-873-17988
; Sequence 17988, Application US/60197873
; GENERAL INFORMATION:
; APPLICANT: Bejanin, Stephane
; APPLICANT: Tanaka, Hiroaki
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Jobert, Severin
; APPLICANT: Giordano, Jean-Yves
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: 81.US1.PRO
; CURRENT APPLICATION NUMBER: US/60/197,873
; CURRENT FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 52153
; SOFTWARE: Patent.pm
; SEQ ID NO 17988
; LENGTH: 128
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -16...-1
; NAME/KEY: UNSURE
; LOCATION: 100
; OTHER INFORMATION: Xaa = Ala, Asp, Gly, Val
US-60-197-873-17988

Query Match 100.0%; Score 62; DB 33; Length 128;
Best Local Similarity 100.0%; Pred. No. 4.9; Length 128;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGHPRPPRGR 10
Db 77 GGHPRPPRGR 86

RESULT 5
US-09-834-366-13405
; Sequence 13405, Application US/09834366
; GENERAL INFORMATION:
; APPLICANT: Bejanin, Stephane
; APPLICANT: Tanaka, Hiroaki
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Jobert, Severin
; APPLICANT: Giordano, Jean-Yves
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: 81.US2.REG
; CURRENT APPLICATION NUMBER: US/09/834,366
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: US 60/197,873
; PRIOR FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 52153
; SOFTWARE: Patent.pm
; SEQ ID NO 13405
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -16...-1
; NAME/KEY: UNSURE
; LOCATION: 33
; OTHER INFORMATION: Xaa = Ala, Pro
; LOCATION: 50
; OTHER INFORMATION: Xaa = Asp, Asn
US-09-834-366-13405

Query Match 100.0%; Score 62; DB 23; Length 131;
Best Local Similarity 100.0%; Pred. No. 5; Length 131;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGHPRPPRGR 10

```

```
Db          115 GGHRPPRGR 124
|||||
RESULT 6
US-60-197-873-13405
; Sequence 13405, Application US/60197873
; GENERAL INFORMATION:
; APPLICANT: Bejanin, Stephane
; APPLICANT: Tanaka, Hiroaki
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Jobert, Severin
; APPLICANT: Giordano, Jean-Yves
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: 81 US1 PRO
; CURRENT APPLICATION NUMBER: US/60/197,873
; CURRENT FILING DATE: 2000-04-18
; NUMBER OF SEQ ID NOS: 52153
; SOFTWARE: Patent.pm
; SEQ ID NO 13405
; LENGTH: 131
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -16...-1
; NAME/KEY: UNSURE
; LOCATION: 33
; OTHER INFORMATION: Xaa = Ala,Pro
; NAME/KEY: UNSURE
; LOCATION: 50
; OTHER INFORMATION: Xaa = Asp,Asn
US-60-197-873-13405
Query Match          100.0%; Score 62; DB 33; Length 131;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 GGHRPPRGR 10
|||||
Db          115 GGHRPPRGR 124
|||||

RESULT 7
PCT-US01-00663-32592
; Sequence 32592, Application PC/TUS0100663
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HUMAN PLACENTA
; FILE REFERENCE: PB 0004 WO 7
; CURRENT APPLICATION NUMBER: PCT/US01/00663
; CURRENT FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 38837
; SOFTWARE: Molecular Dynamics Sequence Listing Engine

PCT-US01-00663-32592
; Sequence 32592, Application US/09864761
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43644
; LENGTH: 132

; SEQ ID NO 32592
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006518.17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EST HUMAN HIT: BF086785.1, EVALUATE 1.00e-01
; OTHER INFORMATION: SWISSPROT HIT: F02810, EVALUATE 7.40e-02
PCT-US01-00663-32592
Query Match          100.0%; Score 62; DB 1; Length 132;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy          1 GGHRPPRGR 10
|||||
Db          81 GGHRPPRGR 90
|||||

RESULT 8
US-09-864-761-43644
; Sequence 43644, Application US/09864761
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43644
; LENGTH: 132
```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006518.17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77
; OTHER INFORMATION: EST HUMAN HIT: BF088785.1, EVALUE 1.00e-01
; OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02
US-09-864-761-43644
Query Match 100.0%; Score 62; DB 23; Length 132;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRPPRGR 10
Db 81 GGHPRPPRGR 90
|||||

RESULT 9
US-10-182-993-31599
; Sequence 31599, Application US/10182993
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: PB 0004 WO 2
; CURRENT APPLICATION NUMBER: US/10/182,993
; PRIOR FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 37811
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 31599
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006518.17
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77
; FEATURE:
; OTHER INFORMATION: EST_HUMAN_HIT: BF088785.1, EVALUE 1.00e-01
; OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02
US-10-182-993-31599
Query Match 100.0%; Score 62; DB 27; Length 132;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRPPRGR 10
Db 81 GGHPRPPRGR 90
|||||

RESULT 10
US-10-203-134-32364
; Sequence 32364, Application US/10203134
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: PB 0004 WO 6
; CURRENT APPLICATION NUMBER: US/10/203,134
; PRIOR FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 38628
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 32364
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006518.17
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
; FEATURE:
; OTHER INFORMATION: EST_HUMAN_HIT: BF088785.1, EVALUE 1.00e-01
; OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02
US-10-203-134-32364
Query Match 100.0%; Score 62; DB 28; Length 132;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRPPRGR 10
Db 81 GGHPRPPRGR 90
|||||

RESULT 11
US-10-203-135-31538
; Sequence 31538, Application US/10203135
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: PB 0004 WO 5
; CURRENT APPLICATION NUMBER: US/10/203,135
; PRIOR FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456

```

```
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 37012
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 31538
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006518.17
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; FEATURE:
; OTHER INFORMATION: EST_HUMAN HIT: BF088785.1, EVALUE 1.00e-01
; FEATURE:
; OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02
; US-10-203-135-31538
Query Match 100.0%; Score 62; DB 28; Length 132;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGHRPPRGR 10
DB 81 GGHRPPRGR 90
RESULT 12
US-10-203-136-32390
; Sequence 32390, Application US/10203136
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HUMAN ADULT LIVER
; FILE REFERENCE: PB 0004 WO 3
; CURRENT APPLICATION NUMBER: US/10/203,136
; PRIOR FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 38837
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 32390
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006518.17
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; FEATURE:
; OTHER INFORMATION: EST_HUMAN HIT: BF088785.1, EVALUE 1.00e-01
; FEATURE:
; OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02
; US-10-203-137-32592
Query Match 100.0%; Score 62; DB 28; Length 132;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGHRPPRGR 10
DB 81 GGHRPPRGR 90
RESULT 13
US-10-203-137-32592
; Sequence 32592, Application US/10203137
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HUMAN PLACENTA
; FILE REFERENCE: PB 0004 WO 7
; CURRENT APPLICATION NUMBER: US/10/203,137
; PRIOR FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 38837
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 32592
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006518.17
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; FEATURE:
; OTHER INFORMATION: EST_HUMAN HIT: BF088785.1, EVALUE 1.00e-01
; FEATURE:
; OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02
; US-10-203-137-32592
Query Match 100.0%; Score 62; DB 28; Length 132;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGHRPPRGR 10
DB 81 GGHRPPRGR 90
RESULT 14
US-10-203-139-31483
; Sequence 31483, Application US/10203139
; GENERAL INFORMATION:
```

```

; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN HUMAN FETAL LIVER
; FILE REFERENCE: PB 0004 WO 4
; CURRENT APPLICATION NUMBER: US/10/203,139
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 37156
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 31483
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006518.17
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; FEATURE:
; OTHER INFORMATION: EST_HUMAN HIT: BF088785.1, EVALUE 1.00e-01
; FEATURE:
; OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02
; US-10-203-139-31483

Query Match 100.0%; Score 62; DB 28; Length 132;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRPGR 10
DB 81 GGHPRPGR 90

RESULT 15
US-10-002-945-24
; GENERAL INFORMATION:
; APPLICANT: Grosche, William
; APPLICANT: MacDougall, John R
; APPLICANT: Smithson, Glenda
; APPLICANT: Millet, Isabelle
; APPLICANT: Stone, David
; APPLICANT: Gunther, Erik
; APPLICANT: Ellerman, Karen
; APPLICANT: Alsbrook II, John P
; APPLICANT: Lepley, Denise M
; APPLICANT: Burgess, Catherine
; APPLICANT: Spytek, Kimberly A
; APPLICANT: Edinger, Shlomit
; APPLICANT: Gangolli, Esha A
; APPLICANT: Gorman, Linda
; APPLICANT: Taupier Jr, Raymond J
; APPLICANT: Li, Li
; APPLICANT: Guo, Xiaojia (Sasha)
; APPLICANT: Fernandes, Elma R
; APPLICANT: Vernet, Corine A M
; APPLICANT: Tchernev, Velizar T
; APPLICANT: Casman, Stacie J
; APPLICANT: Shenoy, Suresh G
; APPLICANT: Mishra, Vishnu
; APPLICANT: Furtak, Katarzyna
; APPLICANT: Baumgartner, Jason C
; APPLICANT: Colman, Steven D
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-206
; CURRENT APPLICATION NUMBER: US/10/002,945
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,291
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: 60/245,317
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: 60/246,562
; PRIOR FILING DATE: 2000-11-07
; PRIOR APPLICATION NUMBER: 60/246,871
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/264,389
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/264,423
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 60/264,799
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 214
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-002-945-24

Query Match 100.0%; Score 62; DB 26; Length 149;
Best Local Similarity 100.0%; Pred. No. 5.6;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRPGR 10
DB 98 GGHPRPGR 107

Search completed: April 6, 2004, 16:57:03
Job time : 208.411 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:56:34 ; Search time 14.6729 Seconds  
(without alignments)  
35.185 Million cell updates/sec

Title: US-10-009-709-7  
Perfect score: 62  
Sequence: 1 GGHPRPPRGR 10

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCUTS COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51	82.3	559	4	US-09-252-991A-24480
2	50	80.6	418	4	US-09-252-991A-29452
3	48	77.4	223	4	US-09-252-991A-17072
4	47	75.8	152	4	US-09-252-991A-31405
5	47	75.8	207	4	US-09-252-991A-29505
6	46	74.2	208	4	US-09-252-991A-32430
7	46	74.2	331	4	US-09-489-039A-11696
8	45	72.6	205	4	US-09-252-991A-25573
9	44	71.0	453	4	US-09-252-991A-27041
10	44	71.0	690	4	US-09-252-991A-16715
11	43	69.4	237	4	US-09-252-991A-16775
12	43	69.4	420	4	US-09-252-991A-30229
13	43	69.4	521	2	US-08-682-847-4
14	42	67.7	18	4	US-09-086-168B-5
15	42	67.7	97	4	US-09-489-039A-8140
16	42	67.7	133	4	US-09-252-991A-32859
17	42	67.7	152	4	US-09-252-991A-18309
18	42	67.7	157	4	US-09-252-991A-23915
19	42	67.7	178	4	US-09-252-991A-16639
20	42	67.7	210	4	US-09-252-991A-24923
21	42	67.7	304	4	US-09-489-039A-8725
22	42	67.7	318	4	US-09-252-991A-23557
23	41	66.1	155	4	US-09-489-039A-12796
24	41	66.1	199	4	US-09-252-991A-22308
25	41	66.1	265	4	US-09-252-991A-21368
26	41	66.1	377	4	US-09-252-991A-29332
27	41	66.1	417	4	US-09-252-991A-18550

28 41 66.1 439 4 US-09-252-991A-16736 Sequence 16736, A  
29 41 66.1 468 4 US-09-485-648-4 Sequence 4, Appli  
30 41 66.1 468 4 US-09-503-565-4 Sequence 4, Appli  
31 41 66.1 468 4 US-09-485-649-4 Sequence 4, Appli  
32 41 66.1 468 4 US-09-333-159B-8 Sequence 8, Appli  
33 41 66.1 493 4 US-09-485-648-2 Sequence 2, Appli  
34 41 66.1 493 4 US-09-503-565-2 Sequence 2, Appli  
35 41 66.1 493 4 US-09-485-649-2 Sequence 2, Appli  
36 41 66.1 493 4 US-09-339-159B-6 Sequence 6, Appli  
37 41 66.1 631 4 US-08-971-188-8 Sequence 8, Appli  
38 41 66.1 676 4 US-09-252-991A-24713 Sequence 24713, A  
39 41 66.1 693 4 US-09-252-991A-26071 Sequence 26071, A  
40 41 66.1 770 4 US-09-252-991A-24516 Sequence 24516, A  
41 66.1 1466 4 US-09-262-537-20 Sequence 20, Appli  
42 41 66.1 1469 4 US-09-262-537-58 Sequence 58, Appli  
43 41 66.1 1471 4 US-08-811-519-1 Sequence 1, Appli  
44 41 66.1 1706 4 US-09-252-991A-31760 Sequence 31760, A  
45 40 64.5 107 4 US-08-849-303-11 Sequence 11, Appli

## ALIGNMENTS

## RESULT 1

US-09-252-991A-24480  
; Sequence 24480, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24480  
; LENGTH: 559  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24480

Query Match 82.3%; Score 51; DB 4; Length 559;  
Best Local Similarity 90.0%; Pred. No. 7.1;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGHPRPPRGR 10  
Db 312 GGLPRPPRGR 321

## RESULT 2

US-09-252-991A-29452  
; Sequence 29452, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 29452  
; LENGTH: 418  
; TYPE: PRT



```
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29452

Query Match      80.6%; Score 50; DB 4; Length 418;
Best Local Similarity 100.0%; Pred. No. 7.6;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HRPFRPRGR 10
Db 160 HRPFRPRGR 167

RESULT 3
US-09-252-991A-17072
; Sequence 17072, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 17072
; LENGTH: 223
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-17072

Query Match      77.4%; Score 48; DB 4; Length 223;
Best Local Similarity 88.9%; Pred. No. 8.1;
Matches 8; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

Qy 2 GHPRPRGR 10
Db 138 GHPRPRGR 146

RESULT 4
US-09-252-991A-31405
; Sequence 31405, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31405
; LENGTH: 152
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31405

Query Match      75.8%; Score 47; DB 4; Length 152;
Best Local Similarity 80.0%; Pred. No. 7.9;
Matches 8; Conservative 0; Mismatches 2; Indels 2; Gaps 0;

Qy 1 GGHPRPRGR 10
Db 77 GAHPRPRGR 86
```

```
RESULT 5
US-09-252-991A-29505
; Sequence 29505, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29505
; LENGTH: 207
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29505

Query Match      75.8%; Score 47; DB 4; Length 207;
Best Local Similarity 88.9%; Pred. No. 10;
Matches 8; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

Qy 2 GHPRPRGR 10
Db 82 GSFRPRGR 90

RESULT 6
US-09-252-991A-32430
; Sequence 32430, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32430
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32430

Query Match      74.2%; Score 46; DB 4; Length 208;
Best Local Similarity 75.0%; Pred. No. 14;
Matches 9; Conservative 0; Mismatches 1; Indels 2; Gaps 1;

Qy 1 GGH--PRPRGR 10
Db 128 GGHARPRPRGR 139

RESULT 7
US-09-489-039A-11696
; Sequence 11696, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
```

; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 11696  
; LENGTH: 331  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-11696

Query Match 74.2%; Score 46; DB 4; Length 331;  
Best Local Similarity 77.8%; Pred. No. 22;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGHPRPPRG 9  
||| |||  
DB 259 GGRPPPKG 267

RESULT 8  
US-09-252-991A-25573  
; Sequence 25573, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 25573  
; LENGTH: 205  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-25573

Query Match 72.6%; Score 45; DB 4; Length 205;  
Best Local Similarity 80.0%; Pred. No. 20;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GGHPRPPRG 10  
||| |||  
DB 91 GGHARHPRG 100

RESULT 9  
US-09-252-991A-27041  
; Sequence 27041, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27041  
; LENGTH: 453  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27041

Query Match 71.0%; Score 44; DB 4; Length 453;

Best Local Similarity 77.8%; Pred. No. 56;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 GHRPPRGR 10  
||| |||  
DB 270 GHPDPGR 278

RESULT 10  
US-09-252-991A-16715  
; Sequence 16715, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 16715  
; LENGTH: 690  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-16715

Query Match 71.0%; Score 44; DB 4; Length 690;  
Best Local Similarity 77.8%; Pred. No. 82;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGHPRPPRG 9  
||| |||  
DB 64 GGYGPPRG 72

RESULT 11  
US-09-252-991A-17675  
; Sequence 17675, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 17675  
; LENGTH: 237  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17675

Query Match 69.4%; Score 43; DB 4; Length 237;  
Best Local Similarity 87.5%; Pred. No. 43;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGHPRPPRG 8  
||| |||  
DB 209 GGRPPRG 216

RESULT 12  
US-09-252-991A-30229  
; Sequence 30229, Application US/09252991A

Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30229  
; LENGTH: 420  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30229

Query Match 69.4%; Score 43; DB 4; Length 420;  
Best Local Similarity 77.8%; Pred. No. 72;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GGHPRPPRG 9  
Db 69 GRHPRPPAG 77

RESULT 13  
US-08-682-847-4  
; Sequence 4, Application US/08682847  
; Patent No. 5858989  
; GENERAL INFORMATION:  
; APPLICANT: BABIUK, LOENE  
; APPLICANT: VAN DEN HURK, SYLVIA  
; APPLICANT: ZAMB, TIM  
; APPLICANT: FITZPATRICK, DAVID  
; TITLE OF INVENTION: RECOMBINANT BOVINE HERPESVIRUS TYPE 1  
; TITLE OF INVENTION: POLYPEPTIDES AND VACCINES  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 PAGE MILL ROAD  
; CITY: PALO ALTO  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/682,847  
; FILING DATE: 12-JUL-1996  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: PARK, FREDDIE K.  
; REGISTRATION NUMBER: 35,636  
; REFERENCE/DOCKET NUMBER: 29310-20005.10  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 813-5600  
; TELEFAX: (415) 494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 521 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-682-847-4

Query Match 69.4%; Score 43; DB 2; Length 521;

Best Local Similarity 87.5%; Pred. No. 87;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Qy 1 GGHPRPPR 8  
Db 121 GGRPRPPR 128

RESULT 14  
US-09-086-168B-5  
; Sequence 5, Application US/09086168B  
; Patent No. 6638941  
; GENERAL INFORMATION:  
; APPLICANT: Castelhamo, Arlindo  
; APPLICANT: Witter, David  
; TITLE OF INVENTION: CONFORMATIONALLY CONSTRAINED PEPTIDOMIMETICS AS BETA-TURN  
; TITLE OF INVENTION: TEMPLATES AND MODULATORS OF SH3 DOMAINS  
; FILE REFERENCE: 60384-B  
; CURRENT APPLICATION NUMBER: US/09/086,168B  
; CURRENT FILING DATE: 1998-05-28  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patent in version 3.1  
; SEQ ID NO 5  
; LENGTH: 18  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: SH3 test peptide  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (1)..(18)  
; OTHER INFORMATION: Synthetic Peptide  
US-09-086-168B-5

Query Match 67.7%; Score 42; DB 4; Length 18;  
Best Local Similarity 87.5%; Pred. No. 5.7;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GGHPRPPR 8  
Db 2 GGRPRPPR 9

RESULT 15  
US-09-489-039A-8140  
; Sequence 8140, Application US/09489039A  
; Patent No. 6610836  
; GENERAL INFORMATION:  
; APPLICANT: Gary Breton et. al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 8140  
; LENGTH: 97  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-8140

Query Match 67.7%; Score 42; DB 4; Length 97;  
Best Local Similarity 87.5%; Pred. No. 26;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 GHRPPRPG 9  
Db 28 GHRPPRPG 35

Search completed: April 6, 2004, 16:19:40

Job time : 14.6729 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:52:34 ; Search time 10.0935 Seconds  
(without alignments)

85.771 Million cell updates/sec

Title: US-10-009-709-8

Perfect score: 56

Sequence: 1 GHPRPGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 78: \*  
1: PIR1: \*  
2: PIR2: \*  
3: PIR3: \*  
4: PIR4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	100.0	166	1	PIHUSC
2	56	100.0	166	2	B25372
3	56	100.0	171	2	A27307
4	44	78.6	561	2	E70610
5	43	76.8	340	2	A24026
6	43	76.8	438	2	B72654
7	42	75.0	448	2	I51893
8	42	75.0	453	2	A35642
9	42	75.0	455	2	S28221
10	41	73.2	550	2	A46419
11	41	73.2	1466	2	T17138
12	41	73.2	1467	2	T18411
13	41	73.2	1471	2	T17149
14	41	73.2	1472	2	T18413
15	41	73.2	1510	2	T17145
16	41	73.2	1515	2	T17156
17	41	73.2	1544	2	EC4743
18	41	73.2	2796	2	JC4743
19	40	71.4	86	2	S00945
20	40	71.4	129	2	F71241
21	40	71.4	185	2	P31020
22	40	71.4	325	2	AC0338
23	40	71.4	326	2	B65037
24	40	71.4	326	2	A91061
25	40	71.4	326	2	G85905
26	40	71.4	326	2	AC0832
27	40	71.4	383	2	S32975
28	40	71.4	691	2	A54741
29	39	69.6	55	2	S29770

## ALIGNMENTS

### RESULT 1

#### PIHUSC

salivary proline-rich phosphoprotein precursor PRH2 [validated] - human

N:Alternate names: salivary acidic proline-rich protein PRH2

N:Contains: peptide P-C (basic proline-rich peptide IB-8b); proline-rich phosphoprotein

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1981 #sequence revision 12-Apr-1996 #text change 08-Dec-2000

C:Accession: A25372; A19803; B57868; A92277; A92254; A94725; A91954; S02564; S02563; J

R:Maeda, N.; Kim, H.-S.; Azen, E.A.; Smithies, O.

J. Biol. Chem. 260, 11123-11130, 1985

A:Title: Differential RNA splicing and post-translational cleavages in the human saliv

A:Reference number: A92492; MUID:85289325; PMID:2993301

A:Accession: A25372

A:Molecule type: mRNA

A:Residues: 1-166 <MAE>

A:Cross-references: GB:K03202; NID:G190481; PIDN:AAA60183.1; PID:G190482

R:Schlesinger, D.H.; Hay, D.I.

Int. J. Pept. Protein Res. 17, 34-41, 1981

A:Title: Primary structure of the active tryptic fragments of human and monkey salivar

A:Reference number: A91757; MUID:81191179; PMID:7228490

A:Accession: A19803

A:Molecule type: protein

A:Residues: 17-46 <SCH>

R:Kim, H.S.; Maeda, N.

J. Biol. Chem. 261, 6712-6718, 1986

A:Title: Structures of two HaeIII-type genes in the human salivary proline-rich protei

A:Reference number: A57868; MUID:86196106; PMID:3009472

A:Accession: B57868

A:Molecule type: DNA

A:Residues: 1-166 <KIM>

A:Cross-references: GB:M13058; NID:G190513; PIDN:AAA98908.1; PID:G190514

R:Wong, R.S.C.; Bennick, A.

J. Biol. Chem. 255, 5943-5948, 1980

A:Title: The primary structure of a salivary calcium-binding proline-rich phosphoprote

A:Reference number: A92277; MUID:80204368; PMID:7380845

A:Contents: protein C

A:Accession: A92277

A:Molecule type: protein

A:Residues: 17-19, 'N', 21-166 <WON>

A:Note: the amino-terminal 46 residues are involved with inhibiting hydroxyapatite for

R:Wong, R.S.C.; Hofmann, T.; Bennick, A.

J. Biol. Chem. 254, 4800-4808, 1979

A:Title: The complete primary structure of a proline-rich phosphoprotein from human sa

A:Reference number: A92254; MUID:79173237; PMID:438215

A:Contents: protein A

A:Accession: A92254

A:Molecule type: protein

A:Residues: 17-19, 'N', 21-122 <WOS>

R:Schlesinger, D.H.; Hay, D.I.

In Peptides: Structure and Biological Function (Proc. Sixth Amer. Peptide Symp.), Gros

A:Title: Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent

A:Reference number: A94425

hypothetical prote  
hypothetical prote  
conserved hypotet  
gene alx3 protein  
hypothetical prote  
heterogeneous nucl  
delta-(1-alpha-ami  
elk1 protein - mou  
nitric triacetate  
calcium channel pr  
hypothetical prote  
PTB-associated spl  
C-arrestin - rat  
hypothetical prote  
hypothetical prote  
protein W03G12.9

30 39 69.6 151 2 H82546  
31 39 69.6 200 2 S55609  
32 39 69.6 226 2 AC0829  
33 39 69.6 346 2 I48185  
34 39 69.6 470 2 C70841  
35 39 69.6 633 2 T02673  
36 39 69.6 3649 1 S18268  
37 38 67.9 429 2 JC4965  
38 38 67.9 450 2 AB3238  
39 38 67.9 1852 2 A37860  
40 38 67.9 2911 2 T20566  
41 37.5 67.0 707 2 A46302  
42 37 66.1 92 2 I70113  
43 37 66.1 107 2 A72701  
44 37 66.1 123 2 D72579  
45 37 66.1 137 2 A88637

salivary proline-rich phosphoprotein precursor PRH1 (allele PIF) - human

C:Species: Homo sapiens (man)  
C:Date: 29-Aug-1987 #sequence revision 29-Aug-1987 #text change 20-Aug-1999  
C:Accession: B25372; A57868; S02562; G38355; S06153; B27307  
R:Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
J. Biol. Chem. 260, 11123-11130, 1985  
A:Title: Differential RNA splicing and post-translational cleavages in the human saliv.  
A:Reference number: A92492; MUID:85289325; PMID:2993301  
A:Accession: B25372  
A:Molecule type: mRNA  
A:Residues: 1-166 <MAR>  
A:Cross-references: GB:K03203; NID:gi90483; PIDN:AAA60184.1; PID:gi90484  
R:Kim, H.S.; Maeda, N.  
J. Biol. Chem. 261, 6712-6718, 1986  
A:Title: Structures of two HaellI-type genes in the human salivary proline-rich protein  
A:Reference number: A57868; MUID:86196105; PMID:3009472  
A:Accession: A57868  
A:Molecule type: DNA  
A:Residues: 1-166 <KM>  
A:Cross-references: GB:M13057; NID:gi190511; PIDN:AAA98807.1; PID:gi190512  
R:Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schlue  
Biochem. J. 255, 15-21, 1988  
A:Title: The primary structures of six human salivary acidic proline-rich proteins (PR  
A:Reference number: S02562; MUID:89061650; PMID:3196309  
A:Accession: S02562  
A:Molecule type: protein  
A:Residues: 47-71 <HAY>  
R:Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
Biochemistry 30, 3351-3356, 1991  
A:Title: Basic proline-rich proteins from human parotid saliva: relationships of the c  
A:Reference number: A38355; MUID:91190684; PMID:1849422  
A:Accession: G38355  
A:Molecule type: protein  
A:Residues: 123-166 <AU>  
R:Robinson, R.; Kaufman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
Biochem. J. 263, 497-503, 1989  
A:Title: Primary structure and possible origin of the non-glycosylated basic proline-r  
A:Reference number: S06153; MUID:90088384; PMID:2688632  
A:Accession: S06153  
A:Molecule type: protein  
A:Residues: 123-166 <ROB>  
R:Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
Am. J. Hum. Genet. 41, 1035-1047, 1987  
A:Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich p  
A:Reference number: A27307; MUID:88074309; PMID:3687941  
A:Contents: allele Pa  
A:Accession: B27307  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 17-41, 'L', 43-118, 'C', 120-166 <AZB>  
A:Cross-references: EMBL:K03203  
C:Genetics:  
A:Gene: GDB:PRH1  
A:Cross-references: GDB:l19515; OMIM:168730  
A:Map position: 12p13.2-12p13.2  
A:Introns: 22/1, 34/1  
C:Superfamily: proline-rich protein  
C:Keywords: phosphoprotein; saliva; tandem repeat

Query Match	100.0%;	Score 56;	DB 1;	Length 166;
Best Local Similarity	100.0%;	Pred. NO. 0.11;		
Matches	9;	Mismatches	0;	Indels 0;
Gaps	0;			

RESULT 3  
A27307

**N; Altern.**

RESULT 2  
B25372

C;Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 29-Aug-1997  
 C;Accession: A27307  
 R;Azer, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A;Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich protein  
 A;Reference number: A27307; MUID:88074309; PMID:3687941  
 A;Accession: A27307  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: DNA  
 A;Residues: 1-171 <AZB>  
 A;Cross-references: EMBL:K03203  
 C;Genetics:  
 A;Gene: GDB:PRH1  
 A;Cross-references: GDB:119515; OMIM:168730  
 A;Map position: 12p13.2-12p13.2  
 C;Superfamily: proline-rich protein  
 C;Keywords: phosphoprotein

Query Match 100.0%; Score 56; DB 2; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 0.11;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHPRPGR 9  
 | | | | | | | | |  
 Db 121 GHPRPGR 129

RESULT 4  
 E70610  
 hypothetical protein Rv1215C - Mycobacterium tuberculosis (strain H37Rv)  
 C;Species: Mycobacterium tuberculosis  
 C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C;Accession: E70610  
 R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
 A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A;Reference number: A70500; MUID:98295987; PMID:9634230  
 A;Accession: E70610  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-561 <COI>  
 A;Cross-references: GB:Z93777; GB:AL123456; NID:93261726; PIDN:CA807817.1; PID:e311160;  
 A;Experimental source: strain H37Rv  
 C;Genetics:  
 A;Gene: Rv1215C

Query Match 78.6%; Score 44; DB 2; Length 561;  
 Best Local Similarity 87.5%; Pred. No. 23;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HPRPGR 9  
 | | | | | | | | |  
 Db 534 YPRPGR 541

RESULT 5  
 A24026  
 erythromycin resistance protein - *Arthrobacter* sp.  
 C;Species: *Arthrobacter* sp.  
 C;Date: 22-Jul-1987 #sequence\_revision 22-Jul-1987 #text\_change 18-Jun-1999  
 C;Accession: A24026  
 R;Roberts, A.N.; Hudson, G.S.; Brenner, S.  
 Gene 35, 259-270, 1985  
 A;Reference number: A24026; MUID:8606275; PMID:4043733  
 A;Accession: A24026  
 A;Molecule type: DNA  
 A;Residues: 1-340 <ROB>  
 A;Cross-references: GB:M11276; NID:9142203; PIDN:AAA22075.1; PID:g142204  
 C;Superfamily: rRNA (adenine-N6)-methyltransferase  
 C;Keywords: antibiotic resistance

Query Match 76.8%; Score 43; DB 2; Length 340;  
 Best Local Similarity 87.5%; Pred. No. 20;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GHPRPGR 8  
 | | | | | | | | |  
 Db 330 GRPRPRG 337

RESULT 6  
 B72654  
 probable histidyl-tRNA synthetase AP50662 - *Aeropyrum pernix* (strain K1)  
 C;Species: *Aeropyrum pernix*  
 C;Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 28-Jul-2000  
 C;Accession: B72654  
 R;Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yanazaki, S.; Haikawa, Y.; Jin-no, K.; Takawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; DNA Res. 6, 83-101, 1999  
 A;Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyrum pernix* strain K1  
 A;Reference number: A72450; MUID:99310339; PMID:10382966  
 A;Accession: B72654  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-438 <KAW>  
 A;Cross-references: DDBJ:AP000060; NID:95104188; PIDN:BAA79634.1; PID:g5104319  
 A;Experimental source: strain K1  
 C;Genetics:  
 A;Gene: AP50662  
 C;Superfamily: histidine-tRNA ligase; histidine-tRNA ligase homology

Query Match 76.8%; Score 43; DB 2; Length 438;  
 Best Local Similarity 87.5%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GHPRPGR 8  
 | | | | | | | | |  
 Db 7 GRPRPRG 14

RESULT 7  
 151883  
 alpha-2B-adrenergic receptor - rat  
 C;Species: *Rattus norvegicus* (Norway rat)  
 C;Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 13-Aug-1999  
 C;Accession: 151883  
 R;Le Josssec, M.; Cloix, J.F.; Pecquery, R.; Giudicelli, Y.; Dause, J.P.  
 Am. J. Hypertens. 8, 177-182, 1995  
 A;Title: Differential sodium regulation between salt-sensitive and salt-resistant Sabr.  
 A;Reference number: 151883; MUID:95275492; PMID:7755946  
 A;Accession: 151883  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-448 <RES>  
 A;Cross-references: EMBL:X74400; NID:9840862; PIDN:CAA52411.1; PID:g940816  
 C;Superfamily: vertebrate rhodopsin  
 C;Keywords: neurotransmitter receptor

Query Match 75.0%; Score 42; DB 2; Length 448;  
 Best Local Similarity 85.7%; Pred. No. 37;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHPRPGR 7  
 | | | | | | | | |  
 Db 249 GHPKPR 255

RESULT 8  
 A35842  
 alpha-2B-adrenergic receptor - rat  
 C;Species: *Rattus norvegicus* (Norway rat)  
 C;Date: 22-Jan-1993 #sequence\_revision 22-Jan-1993 #text\_change 13-Aug-1999  
 C;Accession: A35842

B;Zeng, D.; Harrison, J.K.; D'Angelo, D.D.; Barber, C.M.; Tucker, A.L.; Lu, Z.; Lynch, K.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 3102-3106, 1990  
 A;Title: Molecular characterization of a rat alpha-2B-adrenergic receptor.  
 A;Reference number: A35642; MUID:90222177; PMID:2158103  
 A;Accession: A35642  
 A;Molecule type: mRNA  
 A;Residues: 1-453 <ZEN>  
 A;Cross-references: GB:M32061; NID:g202589; PIDN:AAA40635.1; PID:g202590  
 A;Note: 169-Phe was also found  
 C;Superfamily: vertebrate rhodopsin  
 C;Keywords: G protein-coupled receptor; glycoprotein; transmembrane protein

Query Match 75.0%; Score 42; DB 2; Length 453;  
 Best Local Similarity 85.7%; Pred. No. 37;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GHPRPPR 7  
 DB 254 GHPRPPR 260

RESULT 9  
 S28221  
 C;Species: Mus musculus (house mouse)  
 C;Date: 22-Nov-1993 #sequence\_revision 26-May-1995 #text\_change 13-Aug-1999  
 C;Accession: S28221; JH0693  
 B;Chen, W.M.; Chang, A.C.; Shie, B.J.; Chang, Y.H.; Chang, N.C.A.  
 Biochim. Biophys. Acta 1171, 219-223, 1992  
 A;Title: Molecular cloning and characterization of a mouse alpha(2)C2 adrenoceptor subty  
 A;Reference number: S28221; MUID:93129625; PMID:1336396  
 A;Accession: S28221  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-455 <CHP>  
 A;Cross-references: ENBL:M94583  
 A;Note: the authors translated the codon CCA for residue 161 as Phe, ACT for residue 337  
 R;Chruscinski, A.J.; Link, R.E.; Daunt, D.A.; Barsh, G.S.; Kobalka, B.K.  
 Biochem. Biophys. Res. Commun. 186, 1280-1287, 1992  
 A;Title: Cloning and expression of the mouse homolog of the human alpha2-C2 adrenergic r  
 A;Reference number: JH0693; MUID:92378586; PMID:1354956  
 A;Accession: JH0693  
 A;Molecule type: DNA  
 A;Residues: 6-228,231-455 <CHR>  
 A;Cross-references: GB:L00979; NID:g191547; PIDN:AAA37131.1; PID:g191548  
 C;Superfamily: vertebrate rhodopsin  
 C;Keywords: G protein-coupled receptor; transmembrane protein  
 F;18-43/Domain: transmembrane #status predicted <TM1>  
 F;55-81/Domain: transmembrane #status predicted <TM2>  
 F;90-115/Domain: transmembrane #status predicted <TM3>  
 F;134-158/Domain: transmembrane #status predicted <TM4>  
 F;173-193/Domain: transmembrane #status predicted <TM5>  
 F;377-402/Domain: transmembrane #status predicted <TM6>  
 F;410-435/Domain: transmembrane #status predicted <TM7>

Query Match 75.0%; Score 42; DB 2; Length 455;  
 Best Local Similarity 85.7%; Pred. No. 37;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GHPRPPR 7  
 DB 256 GHPRPPR 262

RESULT 10  
 A46419  
 N;Alternate names: gene XE7 protein  
 C;Species: Homo sapiens (man)  
 C;Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Nov-1999  
 C;Accession: A46419; I68598  
 R;Voland, J.R.; Wyzkowski, R.J.; Huang, M.; Dutton, R.W.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 10425-10429, 1992

A;Title: Cloning and sequencing of a trophoblast-endothelial-activated lymphocyte surf.  
 A;Reference number: A46419; MUID:93066251; PMID:1438229  
 A;Accession: A46419  
 A;Status: preliminary  
 A;Molecule type: nucleic acid  
 A;Residues: 1-550 <VOL>  
 A;Cross-references: GB:M99578; NID:g187241; PIDN:AAA36187.1; PID:g187242  
 A;Experimental source: placenta  
 A;Note: sequence extracted from NCBI backbone (NCBIN:117872, NCBIPI:117873)  
 R;Ellison, J.W.; Ramos, C.; Yen, P.H.; Shapiro, L.J.  
 Hum. Mol. Genet. 1, 691-696, 1992  
 A;Title: Structure and expression of the human pseudoautosomal gene XE7.  
 A;Reference number: I54325; MUID:93258310; PMID:1302606  
 A;Accession: I68598  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-384, 'L' <RES>  
 A;Cross-references: GB:I03426; NID:g340386; PIDN:AAA61303.1; PID:g340387  
 C;Genetics:  
 A;Gene: XE7  
 A;Introns: 254/3; 304/2; 384/3

Query Match 73.2%; Score 41; DB 2; Length 550;  
 Best Local Similarity 77.8%; Pred. No. 64;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GHPRPPRGR 9  
 DB 497 GEPGPRGR 505

RESULT 11  
 T17138  
 C;Species: Rattus norvegicus (Norway rat)  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
 C;Accession: T17138  
 R;Krasnoperov, V.G.; Bittner, M.A.; Beavis, R.; Kuang, Y.; Sainikow, K.V.; Chapurny, O  
 Neuron 18, 925-937, 1997  
 A;Title: Alpha-Latrotoxin stimulates exocytosis by the interaction with a neuronal G-p  
 A;Reference number: Z18710; MUID:97352465; PMID:9208860  
 A;Accession: T17138  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-1466 <KRA>  
 A;Cross-references: ENBL:AF081144; NID:g3695114; PID:g3695115; PIDN:AAC62650.1  
 C;Superfamily: alpha-latrotoxin receptor, calcium-independent

Query Match 73.2%; Score 41; DB 2; Length 1466;  
 Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GHPRPPRGR 9  
 DB 1258 GEPGPRGR 1266

RESULT 12  
 T18411  
 N;Alternate names: brain-specific - bovine  
 N;Alternate names: alpha-latrotoxin receptor, calcium-independent  
 C;Species: Bos primigenius taurus (cattle)  
 C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
 C;Accession: T18411  
 R;Matsushita, H.; Lelianaova, V.G.; Ushkaryov, Y.A.  
 FEBS Lett. 443, 348-352, 1999  
 A;Title: The latrophilin family: multiply spliced G protein-coupled receptors with dif  
 A;Reference number: Z18869; MUID:99148828; PMID:10025961  
 A;Accession: T18411  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA  
 A;Residues: 1-1467 <MAT>  
 A;Cross-references: ENBL:AF111097; NID:g4185801; PID:g4185802; PIDN:AAD09191.1



C;Superfamily: alpha-latrotoxin receptor, calcium-independent  
C;Keywords: G protein-coupled receptor

Query Match 73.2%; Score 41; DB 2; Length 1467;  
Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GHPRPPRGR 9  
| | | | |  
Db 1258 GAPEPPRGR 1266

## RESULT 13

T17149  
CL13A protein - rat

C;Species: Rattus norvegicus (Norway rat)  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C;Accession: T17149

R;Sugita, S.; Ichtchenko, K.; Khvotchev, M.; Sudhof, T.C.  
submitted to the EMBL Data Library, July 1998

A;Description: CL family.

A;Reference number: Z18712

A;Accession: T17149

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1471 <SUG>

A;Cross-references: EMBL:AF081146; NID:g3695118; PID:g3695119; PIDN:AAC62652.1

C;Superfamily: alpha-latrotoxin receptor, calcium-independent

Query Match 73.2%; Score 41; DB 2; Length 1471;  
Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GHPRPPRGR 9  
| | | | |  
Db 1263 GAPEPPRGR 1271

## RESULT 14

T18413

latrophilin-1, brain-specific - bovine

N;Alternate names: alpha-latrotoxin receptor, calcium-independent

C;Species: Bos primigenius taurus (catle)

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C;Accession: T18413

R;Matsushita, H.; Lelianaova, V.G.; Ushkaryov, Y.A.

FEB5 Lett. 443, 348-352, 1999

A;Title: The latrophilin family: multiply spliced G protein-coupled receptors with diff

A;Reference number: Z18869; MUID:99148828; PMID:10025961

A;Accession: T18413

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-1472 <MAT>

A;Cross-references: EMBL:AF111098; NID:g4185803; PID:g4185804; PIDN:AAD09192.1

A;Experimental source: brain

C;Genetics:

A;Gene: LPH1

C;Superfamily: alpha-latrotoxin receptor, calcium-independent

C;Keywords: G protein-coupled receptor

Query Match 73.2%; Score 41; DB 2; Length 1472;  
Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GHPRPPRGR 9  
| | | | |  
Db 1263 GAPEPPRGR 1271

## RESULT 15

T17145

CL1AB protein - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C;Accession: T17145  
R;Sugita, S.; Ichtchenko, K.; Khvotchev, M.; Sudhof, T.C.  
submitted to the EMBL Data Library, July 1998  
A;Description: CL family.  
A;Reference number: Z18712  
A;Accession: T17145  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: mRNA  
A;Residues: 1-1510 <SUG>  
A;Cross-references: EMBL:AF081145; NID:g3695116; PID:g3695117; PIDN:AAC62651.1  
C;Superfamily: alpha-latrotoxin receptor, calcium-independent

Query Match 73.2%; Score 41; DB 2; Length 1510;  
Best Local Similarity 77.8%; Pred. No. 1.7e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GHPRPPRGR 9  
| | | | |  
Db 1302 GAPEPPRGR 1310

Search completed: April 6, 2004, 16:16:53  
Job time: 11.0935 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:31:19 ; Search time 5.80374 Seconds  
(without alignments)  
80.746 Million cell updates/sec

Title: US-10-009-709-8  
Perfect score: 56  
Sequence: 1 GHRPPRGR 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	166	1 PRPC_HUMAN	P02810 homo sapien
2	43	76.8	340	1 ERMA_ARTS3	P09891 arthrobacte
3	43	76.8	438	1 SYH_AERPE	Q9yeb2 aeropyrum p
4	42	75.0	453	1 A2AB_RAT	P19328 rattus norv
5	42	75.0	455	1 A2AB_MOUSE	P30545 mus musculu
6	40	71.4	325	1 RLUD_ECO57	Q8x9f0 escherichia
7	40	71.4	325	1 RLUD_ECOL6	Q8fez9 escherichia
8	40	71.4	325	1 RLUD_ECOLI	P33643 escherichia
9	40	71.4	325	1 RLUD_HAEDU	Q917a7 haemophilus
10	40	71.4	325	1 RLUD_SALT1	Q8xgg2 salmonella
11	40	71.4	325	1 RLUD_YERPE	Q8zdv7 yersinia pe
12	40	71.4	378	1 LFNG_MOUSE	O09010 mus musculu
13	40	71.4	378	1 LFNG_RAT	Q924t4 rattus norv
14	40	71.4	440	1 CU63_MOUSE	P58659 mus musculu
15	39	69.6	343	1 ALX3_HUMAN	O95076 homo sapien
16	39	69.6	343	1 ALX3_MOUSE	O70137 mus musculu
17	39	69.6	576	1 Z384_HUMAN	Q8tf68 homo sapien
18	39	69.6	579	1 Z384_RAT	Q8eqj4 rattus norv
19	39	69.6	618	1 RM24_MOUSE	Q9r0s2 mus musculu
20	39	69.6	633	1 ROR_HUMAN	O43390 homo sapien
21	39	69.6	3649	1 ACV5_NOCLA	P27743 nocardia la
22	38	67.9	196	1 R15E_METAC	Q8tpx0 methanosarc
23	38	67.9	196	1 R15E_METMA	Q8ptu5 methanosarc
24	38	67.9	260	1 DPMI_MOUSE	O70152 mus musculu
25	38	67.9	266	1 DPMI_CRIGR	Q9wu83 cricetus
26	38	67.9	429	1 ELK1_MOUSE	P41969 mus musculu
27	38	67.9	892	1 NAL6_HUMAN	P59044 homo sapien
28	38	67.9	1852	1 CCAS_CYPCA	P22316 cyprinus ca
29	37.5	67.0	707	1 SFPO_HUMAN	P23246 homo sapien
30	37	66.1	92	1 ARRC_RAT	P36576 rattus norv
31	37	66.1	237	1 YMS5_STRCM	Q05071 streptomyce
32	37	66.1	221	1 PRP2_HUMAN	P02812 homo sapien
33	37	66.1	256	1 NEF_FV2RO	P04600 human immun

RESULT 1  
PRPC\_HUMAN STANDARD; PRT; 166 AA.  
AC P02810;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-2)  
DE 3) (PRP-2/PRP-4) (PIF-F/PIF-S) (Protein A/protein C) [Contains:  
DE Peptide P-C].  
GN PRP1 AND PRP2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ALLELES PRH1-4 AND PRH2-1).  
RX MEDLINE=86196106; PubMed=3009472;  
RA Kim H.-S., Maeda N.;  
RT "Structures of two Haell-type genes in the human salivary  
RT proline-rich protein multigene family.";  
RL J. Biol. Chem. 261:6712-6718(1986).  
RN [2]  
RP SEQUENCE FROM N.A. (ALLELES PRH2-1 AND PRH1-PIF).  
RX MEDLINE=85289325; PubMed=293301;  
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
RT "Differential RNA splicing and post-translational cleavages in the  
RT human salivary proline-rich protein gene system.";  
RL J. Biol. Chem. 260:11123-11130(1985).  
RN [3]  
RP SEQUENCE OF 17-166 (PRP-1 TO PRP-4; PIF-F AND PIF-S).  
RX MEDLINE=89061850; PubMed=3196309;  
RA Hay D.I., Bennick A., Schlesinger D.H., Minaguchi K.,  
RA Madapallimattam G., Schluckebier S.K.;  
RT "The primary structures of six human salivary acidic proline-rich  
RT proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-f).";  
RL Biochem. J. 255:15-21(1988).  
RN [4]  
RP SEQUENCE OF 17-166 FROM N.A. (PA ALLELE).  
RX MEDLINE=88074309; PubMed=3887941;  
RA Azen E.A., Kim H.S., Goodman P., Flynn S., Maeda N.;  
RT "Alleles at the PRH1 locus coding for the human salivary-acidic  
RT proline-rich proteins Pa, Db, and PIF.";  
RL Am. J. Hum. Genet. 41:1035-1047(1987).  
RN [5]  
RP SEQUENCE OF 17-166 (PRP-2).  
RX MEDLINE=86222916; PubMed=3710693;  
RA Schlesinger D.H., Hay D.I.;  
RT "Complete covalent structure of a proline-rich phosphoprotein, PRP-2,  
RT an inhibitor of calcium phosphate crystal growth from human parotid  
RT saliva.";  
RL Int. J. Pept. Protein Res. 27:373-379(1986).  
RN [6]  
RP SEQUENCE OF 17-166 (PROTEIN C).  
RX MEDLINE=80204368; PubMed=7380845;  
RA Wong R.S.C., Bennick A.;

## ALIGNMENTS

34 37 66.1 331 1 PRP1\_HUMAN P04280 homo sapien  
35 37 66.1 429 1 GAG\_HTL1A P03345 human t-cel  
36 37 66.1 429 1 GAG\_HTL1C P14076 human t-cel  
37 37 66.1 429 1 GAG\_HTL1M P14077 human t-cel  
38 37 66.1 521 1 VGLC\_HSVBC P14378 bovine hepr  
39 37 66.1 687 1 WRK2\_ARATH Q9fg77 arabidopsis  
40 37 66.1 1385 1 YMS5\_CABEL P14501 caenorhabdi  
41 37 66.1 1596 1 MAM\_DROME P21519 drosophila  
42 37 66.1 1729 1 TABP\_HUMAN Q9c0c2 homo sapien  
43 36 64.3 107 1 CYT2\_ORYSA P20907 oryza sativ  
44 36 64.3 447 1 GASR\_HUMAN P32239 homo sapien  
45 36 64.3 450 1 GASR\_PRANA P30796 praonys nat

"The primary structure of a salivary calcium-binding proline-rich phosphoprotein (protein C), a possible precursor of a related salivary protein A."; J. Biol. Chem. 255:5943-5948(1980).

[7] SEQUENCE OF 17-46 (PROTEIN C).  
MEDLINE=81191179; PubMed=7228490;  
Schlesinger D.H., Hay D.I.;  
"Primary structure of the active tryptic fragments of human and monkey salivary antionic proline-rich proteins."; Int. J. Pept. Protein Res. 17:34-41(1981).

[8] SEQUENCE OF 17-122 (PROTEIN A).  
MEDLINE=79173237; PubMed=438215;  
Wong R.S.C., Hofmann T., Bennick A.;  
"The complete primary structure of a proline-rich phosphoprotein from human saliva."; J. Biol. Chem. 254:4800-4808(1979).

[9] SEQUENCE OF 17-122 (PROTEIN A).  
Schlesinger D.H., Hay D.I.;  
"Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent inhibitor of calcium phosphate precipitation in human parotid saliva."; (In) Gross E., Meienhofer J. (eds.);  
Peptides: structure and biological function (Proceedings of the 6th American peptide symposium), pp.133-136, Pierce Chemical Co., Rockford Il. (1979).

[10] SEQUENCE OF 123-166 (PEPTIDE P-C).  
MEDLINE=80227634; PubMed=7390979;  
Isemura S., Saitoh E., Sanada K.;  
"The amino acid sequence of a salivary proline-rich peptide, P-C, and its relation to a salivary proline-rich phosphoprotein, protein C."; J. Biochem. 87:1071-1077(1980).

[11] VARIANT PRH2-3 LYS-163.  
Azen E.A.;  
"A frequent mutation in the acidic proline-rich protein gene, PRH2, causing a Q14YK change closely adjacent to the bacterial binding domain of the cognate salivary PRP (Pr1) in Afro-Americans."; Hum. Mutat. 13:72-72(1988).

-1- FUNCTION: PRP's act as highly potent inhibitors of crystal growth of calcium phosphates. They provide a protective and reparative environment for dental enamel which is important for the integrity of the teeth.

-1- SUBCELLULAR LOCATION: Secreted.

-1- PTM: Proteolytically cleaved; PRP-2, PRP-1, and PIF-S yield PRP-4, PRP-3 (protein A), and PIF-F, respectively.

-1- POLYMORPHISM: Allele PRH1-4 is also known as PRH1 DA allele; allele PRH2-1 is also known as PR1 or protein C; allele PRH2-3 is also known as PR1'.

-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on use by non-profit institutions as long as its content is in no modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.iesb-sib.ch/announcements> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

-----  
EMBL; K03202; AAA60183.1; -  
EMBL; K03203; AAA60184.1; -  
EMBL; M13057; AAA98807.1; -  
EMBL; M13058; AAA98808.1; -  
Genew; HGNC:9366; PRH1.  
Genew; HGNC:9367; PRH2.  
MIM; 168730; -  
MIM; 168790; -  
MIM; 168710; -  
GO; GO:0005615; C:extracellular space; TAS.  
Repeat; parotid gland; Phosphorylation; Signal; Polymorphism;  
Pyrolydione carboxylic acid.

FT	SIGNAL	1	16
FT	CHAIN	17	166
FT	FT		
FT	CHAIN	17	122
FT	FT		
FT	CHAIN	123	166
FT	DOMAIN	17	46
FT	FT		
FT	MOD_RES	17	17
FT	MOD_RES	24	24
FT	MOD_RES	38	38
FT	VARIANT	20	20
FT	FT		
FT	VARIANT	66	66
FT	FT		
FT	VARIANT	163	163
FT	FT		
FT	CONFLICT	41	41
FT	SEQUENCE	166 AA;	17017 MW; A7DF62BF94E3C3EF CRC64;
FT	SEQ		
Query Match		100.0%;	Score 56; DB 1; Length 166;
Best Local Similarity		100.0%;	Pred. No. 0.041;
Matches	9;	Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1	GHPRPPRGR	9
DB	116	GHPRPPRGR	124
RESULT 2			
ERMA ARTS3			
ID	ERMA ARTS3	STANDARD;	PRT; 340 AA.
AC	P09831;		
DT	01-VAR-1989	(Rel. 10, Created)	
DT	01-VAR-1989	(Rel. 10, Last sequence update)	
DT	10-OCT-2003	(Rel. 42, Last annotation update)	
DE	rRNA adenine N-6-methyltransferase (EC 2.1.1.48) (Macrolide-		
DE	lincosamide-sstreptogramin B resistance protein) (Brythromycin		
DE	resistance protein).		
GN	ERMA		
OS	Arthrobaacter sp. (strain B3381).		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Propionibacterineae; Nocardioidaceae; Aeromicrobium.		
NBI_TaxID=31956;			
RN	[1]		
RP	SEQUENCE FROM N.A.		
EX	MEDLINE=86006275; PubMed=4043733;		
RA	Roberts A.N., Hudson G.S., Brenner S.;		
RT	"An erythromycin-resistance gene from an erythromycin-producing		
RT	strain of Arthrobaacter sp.";		
RL	Gene 35:259-270(1985).		
CC	-I- RESIDUE AT POSITION 2058 IN 23S RNA, RESULTING IN REDUCED		
CC	AFFINITY BETWEEN RIBOSOMES & MACROLIDE-LINCOSAMIDE-STREPTOGRAMIN B		
CC	ANTIBIOTICS.		
CC	-I- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + rRNA = S-adenosyl-L-		
CC	homocysteine + rRNA containing N(6)-methyladenine.		
CC	-I- SIMILARITY: Belongs to the rRNA adenine N-6-methyltransferase		
CC	family.		
CC	CC		

FT	SIGNAL	1	16
FT	CHAIN	17	166
FT	CHAIN	17	122
FT	CHAIN	123	166
FT	DOMAIN	17	46
FT	MOD RES	17	17
FT	MOD_RES	24	24
FT	MOD_RES	38	38
FT	VARIANT	20	20
FT	VARIANT	66	66
FT	VARIANT	163	163
FT	CONFLICT	41	41
FT	SEQUENCE	166 AA;	17017 MW; A7DF62BF94E3C3EF CRC64;
ST	Query Match	100.0%;	Score 56; DB 1; Length 166;
ST	Best Local Similarity	100.0%;	Pred.No. 0.041;
ST	Matches	9; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY		1 GHPRPPRGR 9	
DB		116 GHPRPPRGR 124	
RESULT 2			
ERNA_ARTS3			
ID	ERNA_ARTS3	STANDARD;	FRT; 340 AA.
AC	P09891;		
DT	01-MAR-1989	(Rel. 10, Created)	
DT	01-MAR-1989	(Rel. 10, Last sequence update)	
DE	10-OCT-2003	(Rel. 42, Last annotation update)	
DE	rRNA adenine N-6-methyltransferase (EC 2.1.1.48) (Macrolide-		
DE	lincosamide-streptogramin B resistance protein) (Brythromycin		
DE	resistance protein).		
DE	ERMA.		
GN	Arthroacter sp. (strain B3381).		
OS	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Protonibacteriaceae; Nocardioidaceae; Aeromicrobium.		
OX	NCBI_TaxID=31956;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=86006275; PubMed=4043733;		
RA	Roberts A.N., Hudson G.S., Brenner S.;		
RT	"An erythromycin-resistance gene from an erythromycin-producing		
RT	strain of Arthroacter sp.";		
RL	Gene 35:259-270(1985).		
CC	-1- FUNCTION: THIS PROTEIN PRODUCES A DIMETHYLATION OF THE ADENINE		
CC	RESIDUE AT POSITION 2058 IN 23S RRNA, RESULTING IN REDUCED		
CC	AFFINITY BETWEEN RIBOSOMES & MACROLIDE-LINCOSAMIDE-STREPTOGRAMIN B		
CC	ANTIBIOTICS.		
CC	-1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + rRNA = S-adenosyl-L-		
CC	homocysteine + rRNA containing N(6)-methyladenine.		
CC	-1- SIMILARITY: Belongs to the rRNA adenine N-6-methyltransferase		
CC	family.		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
DR	EMBL; M11276; AAA22075.1; -		
DR	InterPro; IPR001737; RNA_A dimeth.		
DR	InterPro; IPR000051; SAM_bind.		
DR	Pfam; PF00398; RrmAD; 1.		
DR	SMART; SMC0650; rADC; 1.		

DR PROSITE; PS01131; RRNA A DIMETH; 1.  
 KW Antibiotic resistance; Transferase; Methyltransferase.  
 SQ SEQUENCE 340 AA; 37453 MW; 999A714C391952B5 CRC64;

Query Match 76.8%; Score 43; DB 1; Length 340;  
 Best Local Similarity 87.5%; Pred.No. 8.7;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GHPRPRPG 8  
 | | | | |  
 DB 330 GRPRPRPG 337

## RESULT 3

SYH\_AERPE STANDARD; PRT; 438 AA.  
 ID QYEB2;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Histidyl-tRNA synthetase (EC 6.1.1.21) (Histidine--tRNA ligase) (HisRS).  
 GN HISS OR APE0662.  
 OS Aeropyrum pernix.  
 OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;  
 OC Desulfurococaceae; Aeropyrum.  
 OX NCBI\_TaxID=56636;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K1;  
 EX MEDLINE=93310339; PubMed=10382966;  
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,  
 RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,  
 RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,  
 RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,  
 RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,  
 RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic hyper-thermophilic  
 crenarchaeon, Aeropyrum pernix K1.";  
 RL DNA Res. 6:83-101(1999).  
 CC -!- CATALYTIC ACTIVITY: ATP + L-histidyl-tRNA(His).  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AP000060; BAA79634.1; -.  
 CC PIR; B72654; B72654.  
 CC HSWP; O32422; 10E0.  
 CC HAMAP; MF\_00127; -; 1.  
 CC InterPro; IPR004154; HGTP\_anticodon.  
 CC InterPro; IPR004516; HISS.  
 CC InterPro; IPR002314; tRNA-synt\_2b.  
 CC InterPro; IPR006195; tRNA\_ligase\_II.  
 CC Pfam; PF03129; HGTP\_anticodon; 1.  
 CC Pfam; PF00587; tRNA-synt\_2b; 1.  
 CC TIGFAMs; TIGR00442; hisS; 1.  
 CC PROSITE; PS0862; AA\_TRNA\_LIGASE\_II; 1.  
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;  
 KW Complete proteome.  
 SQ SEQUENCE 438 AA; 49111 MW; ECCB67F59A9FB7BC CRC64;

Query Match 76.8%; Score 43; DB 1; Length 438;  
 Best Local Similarity 87.5%; Pred.No. 11;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GHPRPRPG 8  
 | | | | |  
 DB 7 GRPRPRPG 14

## RESULT 4

A2AB\_RAT STANDARD; PRT; 453 AA.  
 ID P19328; Q63021; Q925E4;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Alpha-2B adrenergic receptor (Alpha-2B adrenoceptor).  
 GN ADRA2B.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=90222177; PubMed=2158103;  
 RA Zeng D., Harrison J.K., D'Angelo D.D., Barber C.M., Tucker A.L.,  
 RA Lu Z., Lynch K.R.;  
 RT "Molecular characterization of a rat alpha 2B-adrenergic receptor.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:3102-3106(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley;  
 RA Schaak S., Cussac D., Paris H.;  
 RT "Cloning and characterization of the rat alpha2B-adrenergic receptor  
 RT gene promoter.";  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE OF 6-453 FROM N.A.  
 RC STRAIN=Sabra; TISSUE=Kidney;  
 RX MEDLINE=95275492; PubMed=7755946;  
 RA le Joesec M., Cloix J.F., Pecquery R., Giudicelli Y., Dausse J.P.;  
 RT "Differential sodium regulation between salt-sensitive and salt-  
 RT resistant Sabra rats is not due to any mutation in the renal alpha  
 RT 2B-adrenoceptor gene.";  
 RL Am. J. Hypertens. 8:177-182(1995).  
 CC -!- FUNCTION: Alpha-2 adrenergic receptors mediate the catecholamine-  
 CC induced inhibition of adenylyate cyclase through the action of G  
 CC proteins.

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M32061; AAA40635.1; -.  
 CC EMBL; AF366899; AAKS3388.1; -.  
 CC EMBL; X74400; CAA52411.1; -.  
 CC PIR; A35642; A35642.  
 CC HSWP; P29274; 1MMH.  
 CC InterPro; IPR000276; GPCR\_Rhodpsn.  
 CC Pfam; PF00001; 7tm\_1; 1.  
 CC PRINTS; PR0237; GPCRHOOPS.  
 CC PROSITE; PS00237; G\_PROTEIN\_RECEP\_F1\_1; 1.  
 CC PROSITE; PS0262; G\_PROTEIN\_RECEP\_F1\_2; 1.  
 KW G-protein coupled receptor; Transmembrane; Multigene family;  
 KW Phosphorylation; Lipoprotein; Palmitate.  
 FT DOMAIN 1 17 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 18 42 1 (POTENTIAL).  
 FT DOMAIN 43 54 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 55 80 2 (POTENTIAL).  
 FT DOMAIN 81 90 EXTRACELLULAR (POTENTIAL).

```

FT TRANSMEM 91 113 3 (POTENTIAL).
FT DOMAIN 114 135 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 136 158 4 (POTENTIAL).
FT DOMAIN 159 174 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 175 198 5 (POTENTIAL).
FT DOMAIN 199 375 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 376 399 6 (POTENTIAL).
FT DOMAIN 400 408 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 409 432 7 (POTENTIAL).
FT DOMAIN 433 453 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 454 469 BY SIMILARITY.
FT LIPID 445 445 S-palmitoyl cysteine (Potential).
FT DOMAIN 300 314 ASP/GLU-RICH (ACIDIC).
FT SITE 97 IMPLICATED IN LIGAND BINDING (BY SIMILARITY).
FT SITE 181 IMPLICATED IN CATECHOL AGONIST BINDING (BY SIMILARITY).
FT SITE 185 IMPLICATED IN CATECHOL AGONIST BINDING (BY SIMILARITY).
FT CONFLICT 132 132 R -> C (IN REF. 1).
FT CONFLICT 162 163 EP -> DA (IN REF. 1).
FT SEQUENCE 453 AA; 50369 MW; CBA69CE23EACB511 CRC64;

Query Match 75.0%; Score 42; DB 1; Length 453;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHPRPPR 7
Db 254 GHPRPPR 260

RESULT 5
A2AB MOUSE
ID A2AB MOUSE STANDARD; PRT; 455 AA.
AC P30545;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alpha-2B adrenergic receptor (Alpha-2B adrenoceptor).
GN ADRA2B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DBA/2; TISSUE=Liver;
RX MEDLINE=93129625; PubMed=1336396;
RA Chen W.-M., Chang A.C., Shie B.J., Chang Y.-H., Chang N.-C.A.;
RT "Molecular cloning and characterization of a mouse alpha 2C2
RL adrenoceptor subtype gene."
RL Biochim. Biophys. Acta 1171:219-223(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93379586; PubMed=1354956;
RA Chruscinski A.J., Link R.E., Daunt D.A., Barsh G.S., Kobilka B.K.;
RT "Cloning and expression of the mouse homolog of the human alpha 2-C2
RL adrenergic receptor."
RL Biochem. Biophys. Res. Commun. 186:1280-1287(1992).
CC -!- FUNCTION: Alpha-2 adrenergic receptors mediate the catecholamine-
CC induced inhibition of adenylate cyclase through the action of G
CC proteins.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC ENBL; M94583; AAA73895.1; -
DR ENBL; L00979; AAA37131.1; ALT_INIT.
DR FIR; S28221; S28221.
DR HSP; P29274; 1MMH.
DR MGD; MGI:87935; Adra2b.
DR GO; GO:0004938; F:alpha2-adrenergic receptor activity; IDA.
DR GO; GO:0001525; P:angiogenesis; IMP.
DR GO; GO:0000165; P:MAPKK cascade; IMP.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm_1; 1.
DR PRINTS; PR00237; GPCRHOPOPSN.
DR PROSITE; PS00237; G PROTEIN RECP F1 1; 1.
DR PROSITE; PS0262; G PROTEIN RECP F1 2; 1.
DR G-protein coupled receptor; Transmembrane; Multigene family;
KW Phosphorylation; Lipoprotein; Palmitate
KW DOMAIN 1 17 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 18 42 1 (POTENTIAL).
FT DOMAIN 43 54 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 55 80 2 (POTENTIAL).
FT DOMAIN 81 90 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 91 113 3 (POTENTIAL).
FT DOMAIN 114 135 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 136 158 4 (POTENTIAL).
FT DOMAIN 159 174 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 175 198 5 (POTENTIAL).
FT DOMAIN 199 377 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 378 401 6 (POTENTIAL).
FT DOMAIN 402 410 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 411 434 7 (POTENTIAL).
FT DOMAIN 435 455 CYTOPLASMIC (POTENTIAL).
FT DISULFID 90 169 BY SIMILARITY.
FT LIPID 447 447 S-palmitoyl cysteine (Potential).
FT DOMAIN 306 316 ASP/GLU-RICH (ACIDIC).
FT SITE 97 IMPLICATED IN LIGAND BINDING (BY SIMILARITY).
FT SITE 181 IMPLICATED IN CATECHOL AGONIST BINDING (BY SIMILARITY).
FT SITE 185 IMPLICATED IN CATECHOL AGONIST BINDING (BY SIMILARITY).
FT CONFLICT 202 202 V -> L (IN REF. 1).
FT CONFLICT 229 230 MISSING (IN REF. 2).
FT SEQUENCE 455 AA; 50615 MW; A3954AD76E0E6263 CRC64;

Query Match 75.0%; Score 42; DB 1; Length 455;
Best Local Similarity 85.7%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHPRPPR 7
Db 256 GHPRPPR 262

RESULT 6
RLUD_ECO57
ID RLUD_ECO57 STANDARD; PRT; 325 AA.
AC Q8X9F0;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)
DE (Pseudouridylylate synthase) (Uracil hydrolyase).
GN RLUD OR SPFB OR Z3888 OR ECS3457.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Peria N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

```

RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,  
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
RA Welch R.A., Blattner F.R.;  
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";  
RL Nature 409:529-533(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=O157:H7 / RIMD 0509952;  
RX MEDLINE=21156231; PubMed=11258796;  
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,  
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*  
RT O157:H7 and genomic comparison with a laboratory strain K-12";  
RL DNA Res. 8:11-22(2001).  
RP [2]  
CC -!- FUNCTION: Responsible for synthesis of pseudouridine from uracil  
CC at positions 1911, 1915 and 1917 in 23S ribosomal RNA (By  
CC similarity).  
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine  
CC + H(2)O.  
CC -!- SIMILARITY: Belongs to the pseudouridine synthase rluA family.  
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB005490; AAC57707.1; -;  
CC EMBL; AP002562; BAB36880.1; -;  
CC PIR; A91061; A91061.  
CC InterPro; IPR006225; Pseud synth RluD.  
CC InterPro; IPR006145; Pseudou synth.  
CC InterPro; IPR006224; Rlu synth.  
CC Pfam; PF00849; Pseudou synth\_2; 1.  
CC Pfam; PF01479; S4; 1.  
CC ProDom; PD001819; PSI\_RLU; 1.  
CC SMART; SM00363; S4; 1.  
CC TIGRFAMs; TIGR00005; rluD\_subfam; 1.  
CC PROSITE; PS00889; S4; 1.  
CC InterPro; IPR002942; S4.  
CC INIT MET 0  
CC ACT SITE 138 138 BY SIMILARITY.  
CC ACT SITE 138 138 BY SIMILARITY.  
CC SEQUENCE 325 AA; 36990 MW; 7B046623A4558967 CRC64;  
  
Query Match 71.4%; Score 40; DB 1; Length 325;  
Best Local Similarity 75.0%; Pred. No. 24;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 GHPRPK 8  
DB 254 GRPRPK 261  
  
RESULT 7  
ID RLU1 ECOL6 STANDARD; PRT; 325 AA.  
AC Q8FEZ9; Q83Q10;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)  
DE (Pseudouridylate synthase) (Uracil hydrolyase).  
GN RLU1 OR SFH8 OR C3116 OR SF2653 OR S2829.  
OS *Escherichia coli* O6, and

OS *Shigella flexneri*.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; *Escherichia*.  
OX NCBI\_TaxID=217992, 623;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=E.coli; STRAIN=O6:H1 / CFT073 / ATCC 700928;  
RX MEDLINE=22388234; PubMed=12471157;  
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
RA Rasko D., Buckles E.J., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
RT "Extensive mosaic structure revealed by the complete genome sequence  
RT of uropathogenic *Escherichia coli*";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.flexneri; STRAIN=301 / Serotype 2a;  
RX MEDLINE=2272406; PubMed=12384590;  
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,  
RA Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,  
RA Sun L., Xue Y., Zhao A., Gao Y., Zhu J., Kan B., Ding K., Chen S.,  
RA Cheng H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,  
RA Yu J.;  
RT "Genome sequence of *Shigella flexneri* 2a: insights into pathogenicity  
RT through comparison with genomes of *Escherichia coli* K12 and O157";  
RL Nucleic Acids Res. 30:4432-4441(2002).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.flexneri; STRAIN=2457T / ATCC 700930 / Serotype 2a;  
RX MEDLINE=22590274; PubMed=12704152;  
RA Wei J., Goldberg M.B., Burland V., Venkatesan M.M., Deng W.,  
RA Fournier G., Mayhew G.F., Plunkett G. III, Rose D.J., Darling A.,  
RA Mau B., Perna N.T., Payne S.M., Runyen-Janecky L.J., Zhou S.,  
RA Schwartz D.C., Blattner F.R.;  
RT "Complete genome sequence and comparative genomics of *Shigella*  
RT flexneri serotype 2a strain 2457T";  
RL Infect. Immun. 71:2775-2786(2003).  
CC -!- FUNCTION: Responsible for synthesis of pseudouridine from uracil  
CC at positions 1911, 1915 and 1917 in 23S ribosomal RNA (By  
CC similarity).  
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine  
CC + H(2)O.  
CC -!- SIMILARITY: Belongs to the pseudouridine synthase rluA family.  
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE016764; AAN81565.1; -;  
CC EMBL; AE015281; AAN44149.1; -;  
CC EMBL; AE016987; AAP17972.1; -;  
CC InterPro; IPR006225; Pseud synth RluD.  
CC InterPro; IPR006145; Pseudou synth.  
CC InterPro; IPR006224; Rlu synth.  
CC Pfam; PF00849; Pseudou synth\_2; 1.  
CC Pfam; PF01479; S4; 1.  
CC ProDom; PD001819; PSI\_RLU; 1.  
CC SMART; SM00363; S4; 1.  
CC TIGRFAMs; TIGR00005; rluD\_subfam; 1.  
CC PROSITE; PS01129; PSI\_RLU; 1.  
CC InterPro; IPR00889; S4; 1.  
CC RNA processing; Lyase; RNA-binding; Complete proteome.  
CC INIT MET 0  
CC ACT SITE 17 90 BY SIMILARITY.  
CC ACT SITE 138 138 BY SIMILARITY.  
CC SEQUENCE 325 AA; 36976 MW; 35F58627C17E696C CRC64;

```

RX MEDLINE-20537953; PubMed=11087118;
RA Wrzesinski J., Bakin A., Ofengand J., Lane B.G.;
RT "Isolation and properties of Escherichia coli 23S-RNA pseudouridine
RL IUBMB Life 50:33-37(2000).
RN [7]
RP MUTAGENESIS OF ASP-138.
RX MEDLINE-21345619; PubMed=11453071;
RA Gutgsell N.S., Del Campo M., Raychaudhuri S., Ofengand J.;
RT "A second function for pseudouridine synthases: A point mutant of RluD
RT unable to form pseudouridines 1911, 1915, and 1917 in Escherichia coli
RT 23S ribosomal RNA restores normal growth to an RluD-minus strain.";
RL RNA 7:990-998(2001).
CC -1- FUNCTION: Responsible for synthesis of pseudouridine from uracil
CC at positions 1911, 1915 and 1917 in 23S ribosomal RNA.
CC -1- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC 5'-phosphate + H(2)O.
CC -1- SIMILARITY: Belongs to the pseudouridine synthase rluA family.
CC -1- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -1- CAUTION: Ref.4 sequence differs from that shown due to a
CC frameshift in position 133.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR ENBL; U50134; AAA92957.1; -.
DR ENBL; ABC00346; AAC75643.1; -.
DR ENBL; D90887; BAA16479.1; ALT INT.
DR ENBL; X57620; -. NOT ANNOTATED_CDS.
DR PIR; E65037; E65037.
DR EcoGene; EG12098; rluD.
DR InterPro; IPR006225; Pseud_synth_RluD.
DR InterPro; IPR006145; Pseudou_synth.
DR InterPro; IPR002942; S4.
DR Pfam; PF00849; Pseudou_synth_2; 1.
DR Pfam; PF01479; S4; 1.
DR ProDom; PD001819; PSI_RLU; 1.
DR SMART; SM00363; S4; 1.
DR TIGRFAMs; TIGR00005; rluD_subfam; 1.
DR PROSITE; PS01129; PSI_RLU; 1.
DR PROSITE; PS50889; S4; 1.
KW rRNA processing; Lyase; RNA-binding; Complete proteome.
FT INIT MET 0 0 S4 RNA-BINDING.
FT DOMAIN 17 90
FT ACT_SITE 138 138
FT MUTAGEN 138 138
FT CONFLICT 270 325
FT FT
FT FT
SQ SEQUENCE 325 AA; 36990 MW; 02BDF2AEA4E8300D CRC64;
Query Match 71.4%; Score 40; DB 1; Length 325;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 GHRPPRPG 8
DB 254 GRPRPPKG 261
| | | | |
RESULT 9
ID_RLU_D_ECOLI STANDARD; PRT; 325 AA.
AC P33643; P77003;
DT 01-FEB-1994 (Rel. 28, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)
DE (Pseudouridylylate synthase) (Uracil hydrolyase).
GN RLU_D OR SPFB OR B2594.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Ogura T., Tomoyasu T.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [4]
RP SEQUENCE OF 39-325 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91294165; PubMed=1906060;
RA Kitagawa M., Wada C., Yoshiooka S., Yura T.;
RT "Expression of ClpB, an analog of the ATP-dependent protease
RT regulatory subunit in Escherichia coli, is controlled by a heat shock
RT sigma factor (sigma 32).";
RL J. Bacteriol. 173:4247-4253(1991).
RN [5]
RP CHARACTERIZATION.
RC STRAIN=K12 / MG1655;
RX MEDLINE=99029898; PubMed=9814761;
RA Raychaudhuri S., Conrad J., Hall B.G., Ofengand J.;
RT "A pseudouridine synthase required for the formation of two
RT universally conserved pseudouridines in ribosomal RNA is essential for
RT normal growth of Escherichia coli.";
RL RNA 4:1407-1417(1998).
RN [6]
RP FUNCTION, AND SEQUENCE OF N-TERMINUS.

```

DE Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)  
DE (Pseudouridylylase) (Uracil hydrolyase).  
GN RLUD OR HD0469.  
OS Haemophilus ducreyi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=730;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=3500HP / ATCC 700724;  
RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,  
RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.,  
RA "The complete genome sequence of Haemophilus ducreyi."  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 1-250 FROM N.A.  
RC STRAIN=3500HP / ATCC 700724;  
RX MEDLINE=20200369; PubMed=10735874;  
RA Sun S., Schilling B., Tarantino L., Tullius M.V., Gibson B.W.,  
RA Munson R.S. Jr.;  
RT "Cloning and characterization of the lipooligosaccharide  
RT galactosyltransferase II gene of Haemophilus ducreyi."  
RL J. Bacteriol. 192:2292-2298(2000).  
CC -1- FUNCTION: Responsible for synthesis of pseudouridine from uracil  
CC at positions 1911, 1915 and 1917 in 23S ribosomal RNA (By  
CC similarity).  
CC -1- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine  
CC 5'-phosphate + H(2)O.  
CC -1- SIMILARITY: Belongs to the pseudouridine synthase rluA family.  
CC -1- SIMILARITY: Contains 1 S4 RNA-binding domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; A017152; AAF95428.1; -  
CC EMBL; AF24466; AAF32394.1; -  
CC InterPro; IPR006145; PseudoU synth.  
CC InterPro; IPR006224; Rlu\_synth.  
CC InterPro; IPR002942; S4\_synth.  
CC Pfam; PF00849; PseudoU\_synth\_2; 1.  
CC Pfam; PF01479; S4; 1.  
CC ProDom; PD001819; PSI\_RLU; 1.  
CC SMART; SM00363; S4; 1.  
CC PROSITE; PS01129; PSI\_RLU; 1.  
CC PROSITE; PS00889; S4; 1.  
CC RNA processing; Lyase; RNA-binding; Complete proteome.  
CC DOMAIN 18 78 S4 RNA-BINDING.  
CC FT ACT SITE 139 139 BY SIMILARITY.  
CC SEQUENCE 325 AA; 36990 MW; A93C172718672DE7 CRC64;  
  
Query Match 71.4%; Score 40; DB 1; Length 325;  
Best Local Similarity 75.0%; Pred. No. 24;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
  
Qy 1 GHPRPPRG 8  
Db 255 GRPRPPKG 262  
  
RESULT 10  
RLUD SALTY STANDARD; PRT; 325 AA.  
AC QKXG3;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)  
DE (Pseudouridylylase synthase) (Uracil hydrolyase).

GN OS RLUD OR SFHB OR STM2662 OR STY2851 OR T3618.  
OS Salmonella typhimurium, and  
OS Salmonella typhi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Salmonella.  
OX NCBI\_TaxID=602, 601;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;  
RX MEDLINE=21534948; PubMed=11677609;  
RA McCalland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
RA Courney L., Porwolik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
RA Waterston R., Wilson R.K.;  
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
RT LT2."  
RL Nature 413:852-856(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhi; STRAIN=CT18;  
RX MEDLINE=21534947; PubMed=11677608;  
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,  
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,  
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,  
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,  
RA Feltwell T., Hamlin N., Hague A., Hien T.T., Holroyd S., Jagels K.,  
RA Krog A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,  
RA Quail M.A., Rutherford K., Simmonds M., Skelton J., Stevens K.,  
RA Whitehead S., Barrall B.G.;  
RT "Complete genome sequence of a multiple drug resistant Salmonella  
RT enterica serovar Typhi CT18."  
RL Nature 413:848-852(2001).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC SPECIES=S.typhi; STRAIN=Ty2 / ATCC 700931;  
RX MEDLINE=22531367; PubMed=12644504;  
RA Deng W., Liou S.-R., Plunkett G. III, Mayhew G.F., Rose D.J.,  
RA Burland V., Kodoyanni V., Schwartz D.C., Blattner F.R.;  
RT "Comparative genomics of Salmonella enterica serovar Typhi strains Ty2  
RT and CT18."  
RL J. Bacteriol. 185:2330-2337(2003).  
CC -1- FUNCTION: Responsible for synthesis of pseudouridine from uracil  
CC at positions 1911, 1915 and 1917 in 23S ribosomal RNA (By  
CC similarity).  
CC -1- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine  
CC 5'-phosphate + H(2)O.  
CC -1- SIMILARITY: Belongs to the pseudouridine synthase rluA family.  
CC -1- SIMILARITY: Contains 1 S4 RNA-binding domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; A008821; AAL21552.1; -  
CC EMBL; AL627276; CAD05842.1; -  
CC EMBL; A016842; AAO70189.1; -  
CC StyGene; SG77777; rluD.  
CC InterPro; IPR006225; Pseud\_synth\_RluD.  
CC InterPro; IPR006145; PseudoU synth.  
CC InterPro; IPR006224; Rlu\_synth.  
CC InterPro; IPR002942; S4\_synth.  
CC Pfam; PF00849; PseudoU\_synth\_2; 1.  
CC Pfam; PF01479; S4; 1.  
CC ProDom; PD001819; PSI\_RLU; 1.  
CC SMART; SM00363; S4; 1.  
CC TIGRFAMs; TIGR00005; rluD\_subfam; 1.  
CC PROSITE; PS01129; PSI\_RLU; 1.  
CC PROSITE; PS00889; S4; 1.



```

KW rRNA processing; Lyase; RNA-binding; Complete proteome.
FT INIT_MET 0 BY SIMILARITY.
FT DOMAIN 17 90 S4 RNA-BINDING.
FT ACT_SITE 138 138 BY SIMILARITY.
SQ SEQUENCE 325 AA; 36969 MW; 7AFDD9A1EDF91D9 CRC64;

Query Match 71.4%; Score 40; DB 1; Length 325;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GHPRPPRG 8
DB 254 GRPRPPKG 261

RESULT 11
RLUD_YERPE STANDARD; PRT; 325 AA.
AC Q8ZBV7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)
DE (Pseudouridylate synthase) (uracil hydrolyase).
GN RLUD OR SPFB OR YPO3277 OR Y0912.
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]_
SEQUENCE FROM N.A.
RC STRAIN=CO-92 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Bigham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Feltwell T., Hamlin N., Holroyd S., Jagsels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;
RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL Nature 413:523-527(2001).
RN [2]_
SEQUENCE FROM N.A.
RC STRAIN=KIMS / Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM.";
RL J. Bacteriol. 184:4601-4611(2002).
CC -!- FUNCTION: Responsible for synthesis of pseudouridine from uracil
CC at positions 1911, 1915 and 1917 in 23S ribosomal RNA (By
CC similarity).
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC 5-phosphate + H(2)O.
CC -!- SIMILARITY: Belongs to the pseudouridine synthase rluA family.
CC -!- SIMILARITY: Contains 1 S4 RNA-binding domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ141456; CAC92511.1; -
DR EMBL; AE012693; RAB84494.1; -
DR FIR; AC0398; AC0398.
DR InterPro; IPR006225; Pseud_synth_Rlud.

```

```

DR InterPro; IPR006145; Pseudou synth.
DR InterPro; IPR006224; Rlu_synth.
DR InterPro; IPR002942; S4_
DR Pfam; PF00849; Pseudou_synth_2; 1.
DR Pfam; PF01479; S4; 1.
DR ProDom; PD001819; PSI_RLU; 1.
DR TIGRFAMs; TIGR00005; rluD_subfam; 1.
DR PROSITE; PS01129; PSI_RLU; 1.
DR PROSITE; PS50889; S4_1.
KW rRNA processing; Lyase; RNA-binding; Complete proteome.
FT DOMAIN 18 91 S4 RNA-BINDING.
FT ACT_SITE 139 139 BY SIMILARITY.
SQ SEQUENCE 325 AA; 36706 MW; F05D5B6CDD123ADF CRC64;

Query Match 71.4%; Score 40; DB 1; Length 325;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GHPRPPRG 8
DB 255 GRPRPPKG 262

RESULT 12
LFNG_MOUSE STANDARD; PRT; 378 AA.
AC O9010; Q8K3F1; Q9DC10;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Beta-1,3-N-acetylglucosaminyltransferase lunatic fringe (EC 2.4.1.222)
DE (O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase).
GN LFNG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]_
SEQUENCE FROM N.A.
RX MEDLINE=97330691; PubMed=9187150;
RA Johnston S.H., Rauskolb C., Wilson R., Prabhakaran B., Irvine K.D.,
RA Vogt T.F.;
RT "A family of mammalian fringe genes implicated in boundary
RT determination and the Notch pathway.";
RL Development 124:2245-2254(1997).
RN [2]_
SEQUENCE FROM N.A.
RX MEDLINE=97351513; PubMed=9207795;
RA Cohen B., Bashirullah A., Dagnino L., Campbell C., Fisher W.W.,
RA Leow C.C., Whiting E., Ryan D., Zink D., Boulianne G., Hui C.-C.,
RA Gallie B., Phillips R.A., Lipshitz H.D., Egan S.E.;
RT "Fringe boundaries coincide with Notch-dependent patterning centres in
RT mammals and alter Notch-dependent development in Drosophila.";
RL Nat. Genet. 16:283-288(1997).
RN [3]_
SEQUENCE OF 1-143 FROM N.A., AND DEVELOPMENTAL EXPRESSION.
RC STRAIN=129S6/Syvetac;
RX MEDLINE=22105643; PubMed=12110169;
RA Cole S.E., Levorse J.M., Tilghman S.M., Vogt T.F.;
RT "Clock regulatory elements control cyclic expression of Lunatic fringe
RT during somitogenesis.";
RL Dev. Cell 3:75-84(2002).
RN [4]_
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Lung;
RX MEDLINE=21085660; PubMed=12117851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yananaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Giasi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,

```

RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Wagner L., Washio T.,  
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,  
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,  
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
RA Gustinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuiki S.,  
RA Hayashizaki Y.,  
RA "Functional annotation of a full-length mouse cDNA collection.";  
RL Nature 409:685-690(2001).  
RN [5]  
RP CHARACTERIZATION.  
RX MEDLINE=99264507; PubMed=10330372;  
RA del Barco Barrantes I., Elia A.J., Wuensch K., De Angelis M.H.,  
RA Mak T.W., Rossant J., Conlon R.A., Gossler A., de la Pompa J.L.,  
RT "Interaction between Notch signalling and Lunatic fringe during somite  
boundary formation in the mouse.";  
RL Curr. Biol. 9:470-480(1999).  
RN [6]  
RP CHARACTERIZATION.  
RX MEDLINE=22157274; PubMed=12167404;  
RA Mustonen T., Tuommmers M., Mikami T., Itoh N., Zhang N., Gridley T.,  
RA Thesleff I.,  
RT "Lunatic fringe, FGF, and BMP regulate the Notch pathway during  
epithelial morphogenesis of teeth.";  
RL Dev. Biol. 248:281-293(2002).  
RN [7]  
RP DEVELOPMENTAL EXPRESSION.  
RX MEDLINE=98352783; PubMed=9690473;  
RA Eyraud Y.A., Lun Y., Aulehla A., Gan L., Johnson R.L.,  
RT "Lunatic fringe is an essential mediator of somite segmentation and  
patterning.";  
RL Nature 394:377-381(1998).  
RN [8]  
RP DEVELOPMENTAL EXPRESSION.  
RX MEDLINE=99272806; PubMed=10341080;  
RA Moran J.F., Johnston S.H., Rauekolb C., Bhalerao J., Bowcock A.M.,  
RA Vogt T.F.,  
RT "Genomic structure, mapping, and expression analysis of the mammalian  
Lunatic, Manic, and Radical fringe genes.";  
RL Mamm. Genome 10:535-541(1999).  
RN [9]  
RP CHARACTERIZATION.  
RX MEDLINE=21411855; PubMed=11520458;  
RA Koch U., Lacombe T.A., Holland D., Bowman J.L., Cohen B.L., Egan S.E.,  
RA Guidos C.J.,  
RT "Subversion of the T/B lineage decision in the thymus by lunatic  
fringe-mediated inhibition of Notch-1.";  
RL Immunity 15:225-236(2001).  
RN [10]  
RP DEVELOPMENTAL EXPRESSION.  
RX MEDLINE=22105642; PubMed=12110168;  
RA Morales A.V., Yasuda Y., Ish-Horowitz D.,  
RT "Periodic Lunatic fringe expression is controlled during segmentation  
by a cyclic transcriptional enhancer responsive to notch signaling.";  
RL Dev. Cell 3:63-74(2002).  
RN [11]  
RP FUNCTION: Glycosyltransferase that initiates the elongation of O-  
linked fucose residues attached to EGF-like repeats in the  
extracellular domain of Notch molecules inhibiting Jagged mediated  
Notch signaling. Essential mediator of somite segmentation and  
patterning. During somite boundary formation, it restricts Notch  
activity in the presomitic mesoderm to a boundary-forming territory  
in the posterior half of the prospective somite. In this region,  
Notch function activates a set of genes that are involved in  
boundary formation and in anterior-posterior somite identity.  
Ectopically expressed in the thymus. Lfng inhibits Notch signaling  
which results in inhibition of T cell commitment and promotes B  
cell development in lymphoid progenitors. May play a role in  
boundary formation of the enamel knot.  
CC -/- CATALYTIC ACTIVITY: Transfers a beta-D-GlcNAc residue from UDP-D-

CC GlcNAc to the fucose residue of a fucosylated protein acceptor.  
CC -/- SUBCELLULAR LOCATION: Type II membrane protein. Golgi (By  
similarity).  
CC -/- TISSUE SPECIFICITY: Detected at 12.5 dpc in all tissues examined  
with the highest level observed in adult brain and spleen.  
CC Detected in the dental epithelium.  
CC -/- DEVELOPMENTAL STAGE: Developmental protein. During segmentation it  
shows a cyclic transcription pattern which is under the control of  
Notch. Expressed in the caudal region of the presomitic mesoderm  
with each cycle corresponding to the formation time of one somite.  
CC In the dental epithelium it is detected at stage E13.5. The  
CC pattern of expression corresponds exactly to the formation of the  
CC enamel knot between late bud and early cap stages.  
CC -/- PTM: A soluble form may be derived from the membrane form by  
proteolytic processing.  
CC -/- SIMILARITY: Belongs to the glycosyltransferase family 31.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U94351; AAC53262.1; -;  
CC EMBL; AF015768; AAB71668.1; -;  
CC EMBL; AY124581; AAM91541.1; -;  
CC EMBL; AK004642; -; NOT\_ANNOTATED\_CDS.  
CC GMD; MG1:1095413; Lfng.  
CC GO; GO:0007386; P:compartment specification; IMP.  
CC InterPro; IPR003378; Fringe.  
CC Pfam; PF02434; Fringe; 1.  
CC Signal; Signal: Glycosyltransferase; Developmental protein; Golgi stack;  
CC Signal-anchor; Transmembrane.  
CC DOMAIN 1 8 CYTOPLASMIC (POTENTIAL).  
CC TRANSMEM 9 29 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)  
(POTENTIAL).  
CC DOMAIN 30 378 LUMENAL (POTENTIAL).  
CC SITE 85 86 CLEAVAGE (BY FURIN-LIKE PROTEASE)  
(POTENTIAL).  
CC CARBOHYD 166 166 N-LINKED (GLCNAC... ) (POTENTIAL).  
CC CONFLICT 343 343 V -> M (IN REF. 4).  
CC CONFLICT 361 361 V -> I (IN REF. 4).  
CC SQ SEQUENCE 378 AA; 41952 MW; FDOA02597BF9AFED CRC64;  
Query Match 71.4%; Score 40; DB 1; Length 378;  
Best Local Similarity 100.0%; Pred. No. 28;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GHRPP 6  
DB 100 GHRPP 105  
RESULT 13  
LFNG\_RAT  
ID LFNG\_RAT STANDARD; PRT; 378 AA.  
AC Q924T4;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Beta-1,3-N-acetylglucosaminyltransferase lunatic fringe (EC 2.4.1.222)  
DE [O-fucosylpeptide 3-beta-N-acetylglucosaminyltransferase].  
GN LFNG.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]\_TaxID=10116;  
RP SEQUENCE FROM N.A.  
RA Itoh N.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.



DR HSP; P06601; IFJL.  
 DR TRANSEAC; T04515; -.  
 DR Genew; HGNC:449; ALX3.  
 DR MIM; 606014; -.  
 DR InterPro; IPR001356; Homeobox.  
 DR InterPro; IPR007104; Paired\_homeo.  
 DR Pfam; PF00046; homeobox; 1.  
 DR ProDom; PD000010; Homeobox; 1.  
 DR SMART; SM00389; HOX; 1.  
 DR PROSITE; PS00027; HOMEBOX\_1; 1.  
 DR PROSITE; PS00071; HOMEBOX\_2; 1.  
 KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
 KW Transcription regulation.  
 FT DNA\_BIND 153 212 HOMEBOX.  
 FT CONFLICT 72 72 L -> M (IN REF. 1; AAD01417).  
 FT CONFLICT 86 86 F -> L (IN REF. 1; AAD01417).  
 SQ SEQUENCE 343 AA; 36904 MW; 17AFC7ECD40B942F CRC64;

Query Match 69.6%; Score 39; DB 1; Length 343;  
 Best Local Similarity 85.7%; Pred. No. 37;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HPRPPRG 8  
 |||||  
 Db 39 HPAPPRG 45

Search completed: April 6, 2004, 16:08:00  
 Job time : 6.80374 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:51:34 ; Search time 32.5514 Seconds  
(without alignments)

87.236 Million cell updates/sec

Title: US-10-009-709-8

Perfect score: 56

Sequence: 1 GHPRPRGR 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPTEMBL 25.\*

2: sp\_archaea.\*

3: sp\_bacteria.\*

4: sp\_fungi.\*

5: sp\_human.\*

6: sp\_invertebrate.\*

7: sp\_mammal.\*

8: sp\_muc.\*

9: sp\_organalle.\*

10: sp\_phage.\*

11: sp\_plant.\*

12: sp\_rodent.\*

13: sp\_virus.\*

14: sp\_vertebrate.\*

15: sp\_unclassified.\*

16: sp\_rvirus.\*

17: sp\_bacteriap.\*

18: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	80.4	268	10 Q39487	Q39487 colocasia e
2	45	80.4	407	10 Q8H302	Q8H302 oryza sativ
3	44	78.6	561	16 O05316	O05316 mycobacteri
4	44	78.6	561	16 Q7U0D5	Q7U0D5 mycobacteri
5	43	76.8	545	2 O05089	O05089 nocardioid
6	43	76.8	862	10 Q8W4X6	Q8W4X6 prunus dulc
7	42	75.0	181	2 Q8KEP8	Q8KEP8 synecococc
8	42	75.0	259	16 Q7WIA7	Q7WIA7 bordetella
9	42	75.0	332	2 Q54729	Q54729 synecococc
10	42	75.0	448	11 Q925K7	Q925K7 mus musculu
11	42	75.0	448	11 Q925K6	Q925K6 mus musculu
12	42	75.0	494	5 Q8T777	Q8T777 branchiost
13	42	75.0	535	10 Q84SD0	Q84SD0 oryza sativ
14	42	75.0	707	4 Q8SVG2	Q8SVG2 homo sapien
15	41	73.2	322	11 Q7TNES	Q7TNES mus musculu
16	41	73.2	633	10 Q7XI88	Q7XI88 oryza sativ

17	41	73.2	839	4 Q96IE7	Q96IE7 homo sapien
18	41	73.2	1117	4 Q8NEM6	Q8NEM6 homo sapien
19	41	73.2	1384	2 Q9AJPS	Q9AJPS myxococcus
20	41	73.2	1406	11 Q80TRI	Q80TRI mus musculu
21	41	73.2	1467	6 Q97830	Q97830 bos taurus
22	41	73.2	1469	4 Q9HAR3	Q9HAR3 homo sapien
23	41	73.2	1472	6 Q97831	Q97831 bos taurus
24	41	73.2	1474	4 Q94910	Q94910 homo sapien
25	41	73.2	1515	11 Q88917	Q88917 rattus norv
26	41	73.2	1544	4 Q8WNH8	Q8WNH8 homo sapien
27	41	73.2	2796	2 Q48926	Q48926 mycobacteri
28	40	71.4	74	9 Q94MU2	Q94MU2 bacterioph
29	40	71.4	112	4 Q9H224	Q9H224 homo sapien
30	40	71.4	129	17 Q57931	Q57931 pyrococcus
31	40	71.4	156	12 Q69124	Q69124 human herpe
32	40	71.4	158	12 Q69129	Q69129 human herpe
33	40	71.4	185	16 Q8X2S7	Q8X2S7 escherichia
34	40	71.4	327	12 Q69145	Q69145 human herpe
35	40	71.4	376	2 Q9RHV9	Q9RHV9 corynebacte
36	40	71.4	383	12 Q04397	Q04397 Epstein-Bar
37	40	71.4	383	12 Q8AZK8	Q8AZK8 human herpe
38	40	71.4	383	12 Q8AZK6	Q8AZK6 human herpe
39	40	71.4	383	12 Q8AZK5	Q8AZK5 human herpe
40	40	71.4	383	12 Q8AZK4	Q8AZK4 human herpe
41	40	71.4	383	12 Q8AZK3	Q8AZK3 human herpe
42	40	71.4	383	12 Q8AZK2	Q8AZK2 human herpe
43	40	71.4	383	12 Q8AZK1	Q8AZK1 human herpe
44	40	71.4	383	12 Q8AZK0	Q8AZK0 human herpe
45	40	71.4	437	12 Q69146	Q69146 human herpe

## ALIGNMENTS

### RESULT 1

Q39487 PRELIMINARY; PRT; 268 AA.  
ID Q39487  
AC Q39487;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE 12kb storage protein (Fragment).  
OS Colocasia esculenta (elephant's ear) (Taro).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Araceae; Colocasia.  
OX NCBI\_taxid=4460;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NIVOT ID No.53; TISSUE=Mature tuber;  
RX MEDLINE=94072218; PubMed=8251188;  
RA Hirai M., Nakamura K., Imai T., Sato T.;  
RT "cDNAs encoding for storage proteins in the tubers of taro (Colocasia esculenta Schott).";  
RL Jpn. J. Genet. 68:229-236(1993).  
DR EMBL; D16173; BAA03722.1; -;  
DR HSSP; P30617; IUPC.  
DR GO; GO:0005529; F:sugar binding; IEA.  
DR InterPro; IPR001480; B\_lectin.  
DR Pfam; PF01453; Agglutinin, 2.  
DR SMART; SM00108; B\_lectin; 2.  
FT CHAIN 1 144 268 12-A PROTEIN.  
FT NON\_TER 1 144 268 12-B PROTEIN.  
SQ SEQUENCE 268 AA; 29430 MW; DCDE735578D69FF CRC64;

Query Match 80.4%; Score 45; DB 10; Length 268;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GHPRPR 7

Db 15 GHPRPR 21



DE Hypothetical protein.  
 OS Nocardioides simplex (Arthrobacter simplex).  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Propionibacteriaceae; Nocardioidaceae; Pimeleobacter.  
 OX NCBI\_TaxID=2045;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IFO12069;  
 RX MEDLINE=95319331; PubMed=7596291;  
 RA Molnar I., Choi K., Yamashita M., Murooka Y.;  
 RT "Molecular cloning, expression in Streptomyces lividans, and analysis  
 of a gene cluster from Arthrobacter simplex encoding 3-  
 RT ketosteroid-DEHYA.1-dehydrogenase, 3-ketosteroid-DEHYA.5-isomerase  
 RT and a hypothetical regulatory protein.";  
 RL Mol. Microbiol. 15:895-905(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IFO12069;  
 RX Dzialdek J., Yamashita M., Murooka Y.;  
 RT "Cloning, sequencing and characterization of the downstream region of  
 RT KsdDI operon of Arthrobacter simplex.";  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z93338; CAB07541.1; -;  
 KW Hypothetical protein.  
 SQ SEQUENCE 545 AA; 57318 MW; 55F441EBD5E3A19E CRC64;

Query Match 76.8%; Score 43; DB 2; Length 545;  
 Best Local Similarity 87.5%; Pred. No. 64;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HPRPPRGR 9

Db 456 HPPPPRGR 463

RESULT 6

Q8W4X6  
 ID Q8W4X6 PRELIMINARY; PRT; 862 AA.  
 AC Q8W4X6;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Lipoxigenase (EC 1.13.11.12).  
 GN LOX.  
 OS Prunus dulcis (Almond) (Prunus amygdalus).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.  
 OX NCBI\_TaxID=3755;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=cv. Scorza verde;  
 RA Mita G., Gallo A., Fasano P., Zasiura C., Casey R., Santino A.;  
 RT "Molecular cloning of an almond lipoxigenase gene expressed during the  
 RT early stages of seed development.";  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ418043; CAB10779.2;  
 DR GO; GO:000506; F:iron ion binding; IEA.  
 DR GO; GO:0016165; F:lipoxigenase activity; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR000907; Lipoxigenase.  
 DR InterPro; IPR01024; Lipoxigenase.  
 DR InterPro; IPR008976; FLAT\_LH2.  
 DR Pfam; PF00305; lipoxigenase; 1.  
 DR Pfam; PF01477; FLAT; 1.  
 DR PRINTS; PR00087; LIPOXYGENASE.  
 DR SMART; SW00308; LH2; 1.  
 DR PROSITE; PS00711; LIPOXYGENASE\_1; 1.  
 DR PROSITE; PS00081; LIPOXYGENASE\_2; 1.  
 DR PROSITE; PS50095; FLAT; 1.  
 KW Oxidoreductase.

Qy SEQUENCE 862 AA; 97751 MW; 3FAC9D205DCBC131 CRC64;

Query Match 76.8%; Score 43; DB 10; Length 862;  
 Best Local Similarity 77.8%; Pred. No. 97;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GHPRPPRGR 9

Db 228 GYPYPPRGR 236

RESULT 7

Q8KPP8  
 ID Q8KPP8 PRELIMINARY; PRT; 161 AA.  
 AC Q8KPP8;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE NUGC (EC 1.6.5.3) (NADH dehydrogenase).  
 GN NUGC.  
 OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
 OX NCBI\_TaxID=1140;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PCC 7942;  
 RA Holtman C.K., Sandoval P., Chen Y., Socias T., Mohler B.J.,  
 RA McMurtry S., Gonzalez A., Salinas I., Golden S.S., Youderian P.;  
 RT "Synechococcus elongatus PCC7942 cosmid 7G3.";  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.  
 CC -1- SIMILARITY: BELONGS TO THE COMPLEX I 30 KDA SUBUNIT FAMILY.  
 DR EMBL; AY120853; AAM82722.1; -;  
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.  
 DR InterPro; IPR001268; ComplexI\_30K.  
 DR Pfam; PF00329; complexI\_30Kd; 1.  
 DR ProDom; PD001581; ComplexI\_30K; 1.  
 KW NAD; Oxidoreductase.  
 SQ SEQUENCE 181 AA; 20528 MW; 1021019D31CIAD83 CRC64;

Query Match 75.0%; Score 42; DB 2; Length 181;  
 Best Local Similarity 75.0%; Pred. No. 34;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GHPRPPRGR 8

Db 126 GHPKPPSG 133

RESULT 8

Q7W1A7  
 ID Q7W1A7 PRELIMINARY; PRT; 259 AA.  
 AC Q7W1A7;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN BFP0789.  
 OS Bordetella parapertussis.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Alcaligenaceae; Bordetella.  
 OX NCBI\_TaxID=519;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=12822 / ATCC BAA-587;  
 RX MEDLINE=22827954; PubMed=12910271;  
 RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,  
 RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,  
 RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,  
 RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,  
 RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,  
 RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,

```
RA Leather S., Moule S., Norbaczak H., O'Neil S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640425; CA840198.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 259 AA; 28165 MW; DF32C6FBCA19645 CRC64;

Query Match 75.0%; Score 42; DB 16; Length 259;
Best Local Similarity 77.8%; Pred. No. 47;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GHPRPRGR 9
Db 21 GEPRPTGR 29

RESULT 9
Q54729 PRELIMINARY; PRT; 332 AA.
AC Q925K7;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ORF11692 (Fragment).
OS Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC 7942;
RA Tainoremas N.F., Golden S.S.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U44761; AA86647.1; -.
DR InterPro; IPR001064; Crystallin.
DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
FT NON TER 1
SQ SEQUENCE 332 AA; 38574 MW; CA17B5F4B0F2ED09 CRC64;

Query Match 75.0%; Score 42; DB 2; Length 332;
Best Local Similarity 100.0%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 PRPRGR 9
Db 282 PRPRGR 288

RESULT 10
Q925K7 PRELIMINARY; PRT; 448 AA.
AC Q925K7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Adrenergic receptor alpha 2B.
GN ADRA2B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ISS;
RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;
RT "High-Throughput Sequence Identification of Gene Coding Variants
RT within Alcohol-Related QTLs.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AF332050; AA56079.1; -.
DR MGD; MGI:87935; Adra2b.
DR GO; GO:0004938; F:alpha2-adrenergic receptor activity; IDA.
DR GO; GO:0001525; P:angiogenesis; IMP.
DR GO; GO:0000165; P:MAPKK cascade; IMP.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm.1; 1.
DR PRINTS; PR00237; GPCR_RHODOPSN.
DR PROSITE; PS00237; G PROTEIN RECEPTOR F1.1; 1.
DR PROSITE; PS00262; G PROTEIN RECEPTOR F1.2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
SQ SEQUENCE 448 AA; 49998 MW; B37E5E21B0EC4625 CRC64;

Query Match 75.0%; Score 42; DB 11; Length 448;
Best Local Similarity 85.7%; Pred. No. 77;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHPRPR 7
Db 249 GHPRPR 255

RESULT 11
Q925K6 PRELIMINARY; PRT; 448 AA.
AC Q925K6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Adrenergic receptor alpha 2B.
GN ADRA2B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ILS;
RA Ehringer M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikela J.M.;
RT "High-Throughput Sequence Identification of Gene Coding Variants
RT within Alcohol-Related QTLs.";
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
DR EMBL; AF332050; AA56079.1; -.
DR MGD; MGI:87935; Adra2b.
DR GO; GO:0004938; F:alpha2-adrenergic receptor activity; IDA.
DR GO; GO:0001525; P:angiogenesis; IMP.
DR GO; GO:0000165; P:MAPKK cascade; IMP.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm.1; 1.
DR PRINTS; PR00237; GPCR_RHODOPSN.
DR PROSITE; PS00237; G PROTEIN RECEPTOR F1.1; 1.
DR PROSITE; PS00262; G PROTEIN RECEPTOR F1.2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
SQ SEQUENCE 448 AA; 50018 MW; 1B5ED9456C0B2B73 CRC64;

Query Match 75.0%; Score 42; DB 11; Length 448;
Best Local Similarity 85.7%; Pred. No. 77;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHPRPR 7
Db 249 GHPRPR 255
```

```
CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
DR EMBL; AF332049; AAK5078.1; -.
DR MGD; MGI:87935; Adra2b.
DR GO; GO:0004938; F:alpha2-adrenergic receptor activity; IDA.
DR GO; GO:0001525; P:angiogenesis; IMP.
DR GO; GO:0000165; P:MAPKK cascade; IMP.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR Pfam; PF00001; 7tm.1; 1.
DR PRINTS; PR00237; GPCR_RHODOPSN.
DR PROSITE; PS00237; G PROTEIN RECEPTOR F1.1; 1.
DR PROSITE; PS00262; G PROTEIN RECEPTOR F1.2; 1.
KW G-protein coupled receptor; Receptor; Transmembrane.
SQ SEQUENCE 448 AA; 50018 MW; 1B5ED9456C0B2B73 CRC64;

Query Match 75.0%; Score 42; DB 11; Length 448;
Best Local Similarity 85.7%; Pred. No. 77;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHPRPR 7
Db 249 GHPRPR 255

RESULT 12
Q8T777 PRELIMINARY; PRT; 494 AA.
AC Q8T777;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
```



DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DE Hypothetical protein.  
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).  
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
 CC Branchiostoma.  
 ON NCBI\_TaxID=7739;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=11967531;  
 RA Abi-Rached L., Gilles A., Shiina T., Pontarotti P., Inoko H.;  
 RT "Evidence of an bloc duplication in vertebrate genomes";  
 RL Nat. Genet. 0:0-0(2002).  
 DR EMBL; AF391288; AAM1867.1; .  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR008922; D1-copper centre.  
 DR InterPro; IPR002227; Tyrosinase.  
 DR Pfam; PF00264; tyrosinase.1.  
 DR PRINTS; PRO0092; TYROSINASE.  
 DR PROSITE; PS00497; TYROSINASE\_1; 1.  
 DR PROSITE; PS00498; TYROSINASE\_2; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 494 AA; 56646 MW; 6823A97CDFA105E CRC64;

Query Match 75.0%; Score 42; DB 5; Length 494;  
 Best Local Similarity 87.5%; Pred. No. 85;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 HPRPRGR 9  
 |||||  
 Db 30 HPRSRGR 37

RESULT 13  
 Q84SD0  
 ID Q84SD0 PRELIMINARY; PRT; 535 AA.  
 AC Q84SD0;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DE Hypothetical protein.  
 OS Eukaryota; Viridiplantae; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 CC Ehrhartoideae; Oryzeae; Oryza.  
 ON NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Katayose Y.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, PAC  
 RT clone:P0577B11";  
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP005504; BAC57819.1; .  
 DR InterPro; IPR002885; PPR.  
 DR Pfam; PF01535; PPR; 8.  
 DR TIGRFAMs; TIGR00756; PPR; 9.  
 SQ SEQUENCE 535 AA; 58557 MW; C2712C4178582382 CRC64;

Query Match 75.0%; Score 42; DB 10; Length 535;  
 Best Local Similarity 100.0%; Pred. No. 91;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 PRPRGR 9  
 |||||  
 Db 114 PRPRGR 120

RESULT 14  
 Q86VG2  
 ID Q86VG2 PRELIMINARY; PRT; 707 AA.

AC Q86VG2;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DE Hypothetical protein.  
 OS Branchiostoma floridae (Florida lancelet) (Amphioxus).  
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
 CC Branchiostoma.  
 ON NCBI\_TaxID=7739;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX PubMed=11967531;  
 RA Abi-Rached L., Gilles A., Shiina T., Pontarotti P., Inoko H.;  
 RT "Evidence of an bloc duplication in vertebrate genomes";  
 RL Nat. Genet. 0:0-0(2002).  
 DR EMBL; AF391288; AAM1867.1; .  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR008922; D1-copper centre.  
 DR InterPro; IPR002227; Tyrosinase.  
 DR Pfam; PF00264; tyrosinase.1.  
 DR PRINTS; PRO0092; TYROSINASE.  
 DR PROSITE; PS00497; TYROSINASE\_1; 1.  
 DR PROSITE; PS00498; TYROSINASE\_2; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 494 AA; 56646 MW; 6823A97CDFA105E CRC64;

Query Match 75.0%; Score 42; DB 4; Length 707;  
 Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHPRPR 7  
 |||||  
 Db 229 GHKPRR 235

RESULT 15  
 Q7TNE5  
 ID Q7TNE5 PRELIMINARY; PRT; 322 AA.  
 AC Q7TNE5;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DE Hypothetical protein (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=C57BL/6; Tissue=Brain;  
 MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]

RP SEQUENCE FROM N.A.  
 RX STRAIN=C57BL/6; Tissue=Brain;  
 MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=C57BL/6; Tissue=Brain;  
 MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]

RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL: BC055793; AAH55793.1; --

KW Hypothetical protein.

FT NON\_TER 1

SQ SEQUENCE 322 AA; 33966 MW; 987D23E6FE505774 CRC64;

Query Match 73.2%; Score 41; DB 11; Length 322;

Best Local Similarity 77.8%; Pred. No. 83;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GHPRPRGR 9

Db 114 GGPEPRGR 122

Search completed: April 6, 2004, 16:14:41

Job time : 32.6283 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:29:39 ; Search time 50.3832 Seconds  
(without alignments)  
50.472 Million cell updates/sec

Title: US-10-009-709-8

Perfect score: 56

Sequence: 1 GHRPRPRGR 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_29Jan04.\*  
1: Geneseqp1980s.\*  
2: Geneseqp1980s.\*  
3: Geneseqp2000s.\*  
4: Geneseqp2001s.\*  
5: Geneseqp2002s.\*  
6: Geneseqp2003as.\*  
7: Geneseqp2003bs.\*  
8: Geneseqp2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56	100.0	9	4 AAB48778	Human sal
2	56	100.0	10	4 AAB48777	Human sal
3	56	100.0	17	4 AAB48783	Human sal
4	56	100.0	132	4 AAB38848	Peptide #
5	56	100.0	132	4 AAM32323	Peptide #
6	56	100.0	132	4 AAM72058	Human bon
7	56	100.0	132	4 AAM59494	Human bra
8	56	100.0	132	4 AEG31742	Human liv
9	56	100.0	132	5 AEG41873	Human pep
10	56	100.0	149	6 ABR57423	Human NOV
11	56	100.0	154	6 ABR56769	Human sec
12	56	100.0	166	6 ADA83798	Human PRH
13	56	100.0	166	7 ADC98216	Human sal
14	50	89.3	8	4 AAB48779	Human sal
15	46	82.1	92	6 AAM65798	Propionib
16	46	82.1	141	4 AAU65747	Propionib
17	46	82.1	141	6 ABM62266	Propionib
18	46	82.1	1938	6 ABP76681	Streptomy
19	45	80.4	30	3 AAY65870	Germline
20	45	80.4	843	4 ABG28896	Novel hum
21	44	78.6	51	3 AAG58448	Zea may
22	44	78.6	74	4 AAM23839	Human EST
23	44	78.6	74	4 AAM23850	Human EST
24	44	78.6	74	4 AAM24011	Human EST
25	44	78.6	121	4 AAU49746	Propionib

26	44	78.6	121	6 ABM46265	Propionib
27	42	75.0	7	4 AAB48780	Human sal
28	42	75.0	40	4 ABB42253	Peptide #
29	42	75.0	40	4 AAM36062	Peptide #
30	42	75.0	40	4 AAM75953	Human bon
31	42	75.0	40	4 AAM63140	Human liv
32	42	75.0	40	4 ABG57683	Human liv
33	42	75.0	50	6 ABP80084	N. gonorr
34	42	75.0	51	4 AAU59993	Propionib
35	42	75.0	51	6 ABM56512	Propionib
36	42	75.0	65	4 AAU59580	Propionib
37	42	75.0	65	6 ABM56099	Propionib
38	42	75.0	100	4 AAU45247	Propionib
39	42	75.0	100	6 ABM41766	Propionib
40	42	75.0	104	4 ABG18278	Novel hum
41	42	75.0	210	4 AAU48625	Propionib
42	42	75.0	210	6 ABM45144	Propionib
43	42	75.0	258	4 ABG26281	Novel hum
44	42	75.0	306	4 ABG03818	Novel hum
45	42	75.0	578	7 ADE07941	Novel pro

## ALIGNMENTS

## RESULT 1

AAB48778  
ID AAB48778 standard; peptide; 9 AA.

XX AAB48778;

XX AC

DT 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 100-108), SEQ ID NO:8.

DE Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

KW oral bacterium; caries prevention.

XX Homo sapiens.

OS Homo sapiens.

XX WO200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SE000930.

PF 17-MAY-1999; 99SE-00001773.

PR (STRO//) STROEMBERG N.

XX (JOHA//) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of

proline-rich proteins, useful for preventing dental caries.

PS Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-

CC AAB48783) which contain at least two arginine residues and which protect

CC against dental caries. PRPs (proline-rich proteins) are salivary proteins

CC encoded by six clustered genes on chromosome 12p13.2 and are potential

CC determinants of a person's susceptibility to dental caries. PRPs are

CC degraded by Actinomyces and Streptococcus species to small peptide

CC fragments. These are metabolised by oral bacteria for nutritional

CC purposes, with certain bacterial species generating ammonia via the

CC catabolism of arginine. The peptides of the invention, being arginine-

CC rich, can also be converted to ammonia by these bacteria. The ammonia

CC thus formed raises the pH at the dental surface, thereby protecting the

CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-

CC derived oligopeptides of the invention

XX Sequence 9 AA;

XX Query Match 100.0%; Score 56; DB 4; Length 9;  
XX Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GHPRPPRGR 9

DB 1 GHPRPPRGR 9

RESULT 2

XX AAB48777  
XX ID AAB48777 standard; peptide; 10 AA.

XX AC AAB48777;

XX DT 09-MAR-2001 (first entry)

XX DE Human saliva PRP-1 fragment (residues 99-108), SEQ ID NO:7.

XX KW Human; PRP-1; proline-rich protein; saliva; dental caries;

XX KW Chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

XX KW oral bacterium; caries prevention.

XX OS Homo sapiens.

XX PN WO200069890-A1.

XX PD 23-NOV-2000.

XX PF 11-MAY-2000; 2000WO-SE0000930.

XX PR 17-MAY-1999; 99SE-00001773.

XX PA (STRO/) STROEMBERG N.

XX PA (JOHA/) JOHANSSON I.

XX PI Stroemberg N, Johansson I;

XX DR WPI; 2001-031923/04.

XX PT New oligopeptides comprising 2 arginine residues from degradation of  
XX PT proline-rich proteins, useful for preventing dental caries.

XX PS Claim 4; Page 24; 36pp; English.

XX CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
XX CC AAB48783) which contain at least two arginine residues and which protect  
XX CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
XX CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
XX CC determinants of a person's susceptibility to dental caries. PRPs are  
XX CC degraded by Actinomyces and Streptococcus species to small peptide  
XX CC fragments. These are metabolised by oral bacteria for nutritional  
XX CC purposes, with certain bacterial species generating ammonia via the  
XX CC catabolism of arginine. The peptides of the invention, being arginine-  
XX CC rich, can also be converted to ammonia by these bacteria. The ammonia  
XX CC thus formed raises the pH at the dental surface, thereby protecting the  
XX CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
XX CC derived oligopeptides of the invention

XX Sequence 10 AA;

XX Query Match 100.0%; Score 56; DB 4; Length 10;  
XX Best Local Similarity 100.0%; Pred. No. 0.093;  
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GHPRPPRGR 9

DB 2 GHPRPPRGR 10

RESULT 3

XX AAB48783  
XX ID AAB48783 standard; peptide; 17 AA.

XX AC AAB48783;

XX DT 09-MAR-2001 (first entry)

XX DE Human saliva PRP-1 fragment (residues 99-115), SEQ ID NO:13.

XX KW Human; PRP-1; proline-rich protein; saliva; dental caries;

XX KW Chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

XX KW oral bacterium; caries prevention.

XX OS Homo sapiens.

XX PN WO200069890-A1.

XX PD 23-NOV-2000.

XX PF 11-MAY-2000; 2000WO-SE0000930.

XX PR 17-MAY-1999; 99SE-00001773.

XX PA (STRO/) STROEMBERG N.

XX PA (JOHA/) JOHANSSON I.

XX PI Stroemberg N, Johansson I;

XX DR WPI; 2001-031923/04.

XX PT New oligopeptides comprising 2 arginine residues from degradation of  
XX PT proline-rich proteins, useful for preventing dental caries.

XX PS Claim 2; Page 24; 36pp; English.

XX CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
XX CC AAB48783) which contain at least two arginine residues and which protect  
XX CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
XX CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
XX CC determinants of a person's susceptibility to dental caries. PRPs are  
XX CC degraded by Actinomyces and Streptococcus species to small peptide  
XX CC fragments. These are metabolised by oral bacteria for nutritional  
XX CC purposes, with certain bacterial species generating ammonia via the  
XX CC catabolism of arginine. The peptides of the invention, being arginine-  
XX CC rich, can also be converted to ammonia by these bacteria. The ammonia  
XX CC thus formed raises the pH at the dental surface, thereby protecting the  
XX CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
XX CC derived oligopeptides of the invention

XX Sequence 17 AA;

XX Query Match 100.0%; Score 56; DB 4; Length 17;  
XX Best Local Similarity 100.0%; Pred. No. 0.15;  
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GHPRPPRGR 9

DB 2 GHPRPPRGR 10

RESULT 4

XX ABB38848  
XX ID ABB38848 standard; peptide; 132 AA.

XX AC ABB38848;

XX DT 04-FEB-2002 (first entry)

XX DE Peptide #6354 encoded by human foetal liver single exon probe.

XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.

```

XX OS Homo sapiens.
XX FN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000669.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human fetal liver.
XX Claim 27; SEQ ID NO 31483; 639pp + Sequence Listing; English.
XX The invention relates to a single exon nucleic acid probe for measuring
XX human gene expression in a sample derived from human foetal liver. The
XX single exon nucleic acid probes may be used for predicting, measuring and
XX displaying gene expression in samples derived from human fetal liver. The
XX present sequence is a peptide encoded by a single exon nucleic acid probe
XX of the invention. Note: the sequence data for this patent did not form
XX part of the printed specification, but was obtained in electronic format
XX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX Sequence 132 AA;
XX Query Match 100.0%; Score 56; DB 4; Length 132;
XX Best Local Similarity 100.0%; Pred. No. 0.97;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 GHPRPGR 9
XX DB 82 GHPRPGR 90
XX RESULT 5
XX ID AAM32323 standard; protein; 132 AA.
XX AC AAM32323;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #6360 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000663.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.

```

```

PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human placenta.
XX Claim 27; SEQ ID NO 32592; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENPs;
XX see AA131315-AA157546). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders
XX Sequence 132 AA;
XX Query Match 100.0%; Score 56; DB 4; Length 132;
XX Best Local Similarity 100.0%; Pred. No. 0.97;
XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX QY 1 GHPRPGR 9
XX DB 82 GHPRPGR 90
XX RESULT 6
XX ID AAM72058 standard; protein; 132 AA.
XX AC AAM72058;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 32364.
XX KW Human; bone marrow expressed exon; gene expression analysis; probe;
XX microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US000668.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX Example 4; SEQ ID NO 32364; 658pp + Sequence Listing; English.
XX

```

CC The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention  
 SQ Sequence 132 AA;

Query Match 100.0%; Score 56; DB 4; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 0.97; 0; Indels 0; Gaps 0;  
 Matches 9; Conservative 0; Mismatches 0;

Qy 1 GHPRPGR 9  
 Db 82 GHPRPGR 90

RESULT 7  
 AAM59494  
 ID AAM59494 standard; protein; 132 AA.

XX AC AAM59494;  
 XX DT 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe encoded protein SEQ ID NO: 31599.  
 DE Human; brain expressed exon; gene expression analysis; probe; microarray;  
 KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.  
 XX Homo sapiens.

OS  
 XX WO200157275-A2.  
 PN 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000667.

XX 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human  
 PT brains.

XX Example 4; SEQ ID NO 31599; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC brain. They can be used to measure gene expression in brain cell samples,  
 CC which may enable the diagnosis and improved treatment of nervous system  
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,  
 CC epilepsy and cancers. The present sequence is a protein encoded by one of  
 CC the probes of the invention

SQ Sequence 132 AA;

Query Match 100.0%; Score 56; DB 4; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 0.97; 0; Indels 0; Gaps 0;  
 Matches 9; Conservative 0; Mismatches 0;

Qy 1 GHPRPGR 9

Db 82 GHPRPGR 90

RESULT 8  
 ABG53742  
 ID ABG53742 standard; peptide; 132 AA.

XX AC ABG53742;  
 XX DT 25-FEB-2003 (first entry)

XX Human liver peptide, SEQ ID No 32390.

XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
 KW hypercholesterolaemia; coronary heart disease.

XX Homo sapiens.

XX WO200157273-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000664.

XX 04-FEB-2000; 2000US-0180312P.  
 PR 26-MAY-2000; 2000US-0207456P.  
 PR 30-JUN-2000; 2000US-00608408.  
 PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 XX WPI; 2001-488898/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human adult liver.

XX Claim 27; SEQ ID NO 32390; 658pp; English.

XX The invention relates to a single exon nucleic acid probe (SEN) (I) for  
 CC measuring human gene expression in a sample derived from human adult  
 CC liver, comprising one of 13109 defined nucleotide sequences given in the  
 CC specification (or complements/ fragments). The probe hybridises at high  
 CC stringency to a nucleic acid molecule expressed in the human adult liver.  
 CC (I) may be used for predicting, measuring and displaying gene expression  
 CC in samples derived from human adult liver. The genes identified may be  
 CC involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
 CC associated with coronary heart disease. ABG47348-ABG5930 represent human  
 CC liver single exon encoded peptides of the invention. Note: The sequence  
 CC information for this patent does not appear in the printed specification  
 CC but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 132 AA;

Query Match 100.0%; Score 56; DB 4; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 0.97;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHPRPGR 9  
 Db 82 GHPRPGR 90

RESULT 9  
 ABG41873  
 ID ABG41873 standard; peptide; 132 AA.



PS Claim 1; Page 69; 346pp; English.

XX ACFO3547 to ACF03570 encode the human NOVX proteins (I) given in ABR57412  
 XX to ABR57435. (I) have cytostatic, cardiant, antiinflammatory, neurotropic,  
 CC immunosuppressive, antiallergic, haemostatic, anti-HIV, antidiabetic,  
 CC antiarteriosclerotic, anorectic, antiasthmatic, nephrotropic, virucide,  
 CC antiarthritic, hepatotropic, neuroprotective, antibacterial, relaxant,  
 CC antiparasitic, anticonvulsant, neuroprotective, vasotropic, antiparkinsonian,  
 CC vulnerary, angiogenic and antiangiogenic activities, and can be used in  
 CC gene therapy and vaccines. The NOVX polypeptides and their antibodies can  
 CC be used to determine the presence or absence of (I) in a sample. The NOVX  
 CC polypeptides, polynucleotides encoding them, and antibodies against them,  
 CC are useful in manufacturing a medicament for treating or preventing a  
 CC syndrome associated with a NOVX-associated disorder such as hypertension,  
 CC cardiomyopathy, atherosclerosis, cancer, diabetes, asthma, inflammation,  
 CC autoimmune disorders, allergies, blood disorders, obesity, acquired  
 CC immunodeficiency syndrome (AIDS), immunoglobulin (Ig)A nephropathy,  
 CC cirrhosis, arthritis, Alzheimer's disease, Parkinson's disease, goitre,  
 CC infections (e.g. bacterial, viral, parasitic), stroke, muscular  
 CC dystrophy, epilepsy, and other wasting disorders associated with chronic  
 CC diseases. ACF03571 to ACF03644 represent PCR primers and probes for NOVX  
 CC sequence, which are used in an example from the present invention

XX SQ Sequence 149 AA;

Query Match 100.0%; Score 56; DB 6; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GHPRPGR 9  
 |||||  
 Db 99 GHPRPGR 107

RESULT 11  
 ABR56769  
 ID ABR56769 standard; protein; 154 AA.  
 XX AC ABR56769;  
 XX AC ABR56769;  
 XX DT 30-JUL-2003 (first entry)  
 XX DE Human secreted protein SECP-44 SEQ ID NO:44.  
 XX KW Human; secreted protein; SECP; cytostatic; antiarteriosclerotic;  
 KW anticonvulsant; neurotropic; neuroprotective; cerebroprotective; anti-HIV;  
 KW antiallergic; antiinflammatory; thyromimetic; gene therapy; cancer;  
 KW cell proliferative disorder; atherosclerosis; neurological disorder;  
 KW epilepsy; Huntington's disease; stroke; immune disorder; allergy; AIDS;  
 KW inflammatory disorder; developmental disorder; hypothyroidism;  
 KW Cushing's syndrome; infection.

XX OS Homo sapiens.  
 XX FN WO2003016506-A2.  
 XX PD 27-FEB-2003.  
 XX PF 15-AUG-2002; 2002WO-US027143.  
 XX PR 17-AUG-2001; 2001US-0313249P.  
 XX PR 24-AUG-2001; 2001US-0314752P.  
 XX PR 07-SEP-2001; 2001US-0317818P.  
 XX PR 07-SEP-2001; 2001US-0317824P.  
 XX PR 21-SEP-2001; 2001US-0324040P.  
 XX PR 24-SEP-2001; 2001US-0324586P.  
 XX PR 02-NOV-2001; 2001US-0343980P.  
 XX PR 28-NOV-2001; 2001US-0334229P.  
 XX PR 13-FEB-2002; 2002US-0357002P.  
 XX PR 06-MAR-2002; 2002US-0362439P.  
 XX PR 19-MAR-2002; 2002US-0366041P.  
 XX PR 30-APR-2002; 2002US-0376988P.

XX PA (INCY-) INCYTE GENOMICS INC.

XX PI Tang TY, Warren BA, Gietzen KJ, Lal PG, Yue H, Honchell CD;  
 PI Lehr-Mason PM, Burford N, Xu Y, Baughn MR, Duggan BM, Tran UK;  
 PI Lee RA, Forsythe IJ, Richardson TW, Lee S, Thangavelu K, Yue H;  
 PI Emerling BM, Wallia NK, Azimzai Y, Sanjanwala B, Hafalia AJA;  
 PI Borowsky ML, Nguyen DB, Leon CH, Astromoff A, Ding L, Lee SY;  
 PI Becha SD, Ramkumar J, Gandhi AR, Jin P, Fu GK, Swarnakar A;  
 XX WPI; 2003-278569/27.  
 DR N-PSDB; ACC79069.

XX New human secreted proteins (SECP), useful for diagnosing, treating and  
 PT preventing diseases or conditions associated with the aberrant SECP  
 PT expression e.g. cancer, AIDS, atherosclerosis, Huntington's disease,  
 PT stroke, infections.

XX Claim 1; Page 222; 286pp; English.

PS ACCT79026 to ACR79105 encode the human secreted proteins (I) given in  
 CC ABR5726 to ABR56805, designated SECP-1 to SECP-80. SECP sequences can  
 CC have cytostatic, antiarteriosclerotic, anticonvulsant, antiinflammatory,  
 CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic and  
 CC thymimetic activities, and can be used in gene therapy. The SECP  
 CC proteins and polynucleotides can be used in diagnosing, treating and  
 CC preventing diseases or conditions associated with the decreased  
 CC expression or overexpression of SECP, such as cell proliferative (e.g.  
 CC cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's  
 CC disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and  
 CC developmental (e.g. hypothyroidism, Cushing's syndrome) disorders, or  
 CC infections. They are also useful in assessing the effects of exogenous  
 CC compounds on the expression of nucleic acid and amino acid sequences of  
 CC SECP. The SECP or its fragments are useful in screening compounds for  
 CC effectiveness as agonist or antagonist of the polypeptides, or in  
 CC altering the expression of the target polynucleotide and compounds that  
 CC specifically bind to or modulate the activity of the polypeptide

XX SQ Sequence 154 AA;

Query Match 100.0%; Score 56; DB 6; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 1.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GHPRPGR 9  
 |||||  
 Db 104 GHPRPGR 112

RESULT 12  
 ADA83798  
 ID ADA83798 standard; protein; 166 AA.  
 XX AC ADA83798;  
 XX DT 20-NOV-2003 (first entry)  
 XX DE Human PRH2 protein.  
 XX KW human; marker; expressed sequence tag; EST; arabidopsis; tumour;  
 KW stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;  
 KW vaccine.  
 XX OS Homo sapiens.  
 XX FN WC2002103028-A2.  
 XX PD 27-DEC-2002.  
 XX PR 30-MAY-2002; 2002WO-IB004189.  
 XX PR 30-MAY-2001; 2001US-0291999P.  
 XX PR 22-OCT-2001; 2001US-0330457P.



PR 19-FEB-2002; 2002US-0357144P.  
 PA (BIOM-) BIOMEDICAL CENT.  
 XX Baranova AV, Yankovsky NK, Kozlov AP, Lobashev AV, Krukovskaya LL;  
 XX N-PSDB; ADA83797.  
 DR WPI; 2003-175241/17.  
 XX Determining if a nucleic acid is a marker for a phenotype/cell type of  
 PT interest, by global comparison of expressed sequence tags known to be  
 PT expressed in the phenotype/cell type with all ESTs expressed in normal  
 PT tissue.  
 XX Claim 29; Page 191-192; 516pp; English.  
 PS The invention relates to a novel method for determining if a nucleic acid  
 CC is a marker for a predetermined phenotype/cell type of interest from a  
 CC biological species. The method comprises performing a global comparison  
 CC of a group of expressed sequence tags (ESTs) known to be expressed in the  
 CC phenotype/cell type of interest with all ESTs expressed in normal tissue  
 CC in order to identify ESTs that are preferentially expressed in the  
 CC phenotype/cell of interest. A method of the invention is useful for  
 CC determining whether a nucleic acid is a marker for a predetermined  
 CC phenotype or cell type of interest from a biological species, preferably  
 CC Arabidopsis or human. The cell type of interest is an abnormal cell such  
 CC as a tumour cell, and the predetermined phenotype is a stress-induced  
 CC phenotype such as hyperosmotic stress or high salt conditions. A method  
 CC of the invention is also useful for determining the progression of colon  
 CC cancer in a human, for detecting a tumour cell, and for regulating or  
 CC preventing the growth of a tumour cell. An antibody of the invention is  
 CC useful for detecting the absence or presence of peptides encoded by  
 CC tumour-associated markers. A polypeptide of the invention is useful as an  
 CC immunogen for vaccinating an animal. The present sequence represents a  
 CC tumour-associated antigen of the invention.  
 XX Sequence 166 AA;  
 SQ

Query Match 100.0%; Score 56; DB 6; Length 166;  
 Best Local Similarity 100.0%; Pred. NO. 1.2;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GHPRPPRGR 9  
 DB 116 GHPRPPRGR 124

RESULT 13  
 ADC98216  
 ID ADC98216 standard; protein; 166 AA.  
 AC ADC98216;  
 XX 01-JAN-2004 (first entry)  
 DT Human salivary acidic proline-rich phosphoprotein (PRP).  
 DE Human; salivary acidic proline-rich phosphoprotein; PRP; autoantigen;  
 KW atopic dermatitis-inducing protein; salivary gland; IGE autoantibody;  
 KW immunoglobulin E; mast cell activation; basophil activation; diagnosis;  
 KW risk assessment; sensitisation remedy; dermatological; anti-allergic;  
 KW anti-inflammatory.  
 XX Homo sapiens.  
 OS  
 XX WO2003084991-A1.  
 PN 16-OCT-2003.  
 PD 04-APR-2003; 2003WO-JP004325.  
 PF 08-APR-2002; 2002JP-00105425.  
 PR New oligopeptides comprising 2 arginine residues from degradation of

PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.  
 XX Hide M, Yamamoto S, Tanaka T, Koro O;  
 XX WPI; 2003-833567/77.  
 DR N-PSDB; ADC98215.  
 XX Atopic dermatitis-inducing proteins, applicable in diagnosis of including  
 PT risk of onset, and in developing sensitization remedies for the disease.  
 PT Claim 4; SEQ ID NO 2; 43pp; Japanese.  
 PS The invention relates to the human atopic dermatitis-inducing proteins,  
 CC salivary acidic proline-rich phosphoprotein (PRP; ADC98216) and prolactin  
 CC -inducible protein (PIP; ADC98218), and their post-translationally  
 CC modified forms. These proteins are secreted by salivary or sweat glands  
 CC and bind to IGE autoantibodies, thereby activating mast cells and  
 CC basophils. The invention also relates to antigenic peptide fragments of  
 CC PRP or PIP; antibodies which bind to PRP or PIP; methods for diagnosing  
 CC atopic dermatitis or for determining whether an individual is at risk of  
 CC developing atopic dermatitis by determining the presence of PRP- or PIP-  
 CC specific antibodies or immune complexes, or by quantifying histamine  
 CC release; and sensitization remedies for atopic dermatitis containing PRP  
 CC and/or PIP or their peptides as the active ingredient. PRP, PIP and their  
 CC antibodies are useful in diagnosing atopic dermatitis, or for determining  
 CC whether an individual is at risk of developing atopic dermatitis. They  
 CC are also useful in developing sensitisation remedies for the treatment of  
 CC atopic dermatitis. The present sequence represents the specifically  
 CC claimed human salivary acidic proline-rich phosphoprotein (PRP).  
 XX Sequence 166 AA;  
 SQ

Query Match 100.0%; Score 56; DB 7; Length 166;  
 Best Local Similarity 100.0%; Pred. NO. 1.2;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GHPRPPRGR 9  
 DB 116 GHPRPPRGR 124

RESULT 14  
 AAB48779  
 ID AAB48779 standard; peptide; 8 AA.  
 XX AAB48779;  
 AC 09-MAR-2001 (first entry)  
 DT Human saliva PRP-1 fragment (residues 101-108), SEQ ID NO:9.  
 DE Human; PRP-1; proline-rich protein; saliva; dental caries;  
 KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.  
 XX Homo sapiens.  
 OS  
 XX WO200069890-A1.  
 PN 23-NOV-2000.  
 PD 11-MAY-2000; 2000WO-SE000930.  
 PF 17-MAY-1999; 99SE-00001773.  
 PR (STRO/) STROEMBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX Stroemberg N, Johansson I;  
 PI WPI; 2001-031923/04.  
 DR New oligopeptides comprising 2 arginine residues from degradation of

PT proline-rich proteins, useful for preventing dental caries.

PS Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-48783 represent the PRP-1-  
 CC derived oligopeptides of the invention

XX Sequence 8 AA;

SQ Query Match 89.3%; Score 50; DB 4; Length 8;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 HRP RPGR 9  
 DB 1 HRP RPGR 8

# RESULT 15

ABM65798  
 ID ABM65798 standard; protein; 92 AA.

XX AC ABM65798;

DT 20-OCT-2003 (first entry)

XX Propionibacterium acnes immunogenic polypeptide #30474.

XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
 KW immunostimulant; immune response; vaccine; immunogenic.

XX OS Propionibacterium acnes.

XX PN WO2003033515-A1.

XX PD 24-APR-2003.

XX PF 11-OCT-2002; 2002WO-US032727.

XX PR 15-OCT-2001; 2001US-00978825.

XX PA (CORI-) CORIXA CORP.

XX PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
 PI Barth B, Valliave-Douglass J;

XX DR WPI; 2003-381789/36.

XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.

PS Claim 7; SEQ ID NO 30474; 1481pp; English.

XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 CC encoding a Propionibacterium acnes protein. The invention also relates to  
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC invention; fusion proteins comprising a polypeptide of the invention; a

CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising T cells prepared  
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a  
 CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridisation. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a specifically claimed P. acnes polypeptide which is  
 CC thought to contain an immunogenic region. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

SQ Sequence 92 AA;

Query Match 82.1%; Score 46; DB 6; Length 92;  
 Best Local Similarity 88.9%; Pred. No. 19;  
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GHP RPGR 9  
 DB 15 GHP RPGR 23

Search completed: April 6, 2004, 16:06:39  
 Job time : 50.3832 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 16:14:50 ; Search time 35.3271 Seconds  
(without alignments)  
66.905 Million cell updates/sec

Title: US-10-009-709-8  
Perfect score: 56  
Sequence: 1 GHPRRPRGR 9

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 262633353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Query Match	Score	Length	DB ID	Description
1	56	100.0	132	9	US-09-864-761-43644
2	56	100.0	166	14	US-10-157-031-80
3	46	82.1	19652	15	US-10-084-846A-7
4	45	80.4	238	12	US-10-425-114-69167
5	45	80.4	277	12	US-10-425-114-43478
6	42	75.0	40	9	US-09-864-761-45389
7	42	75.0	72	12	US-10-424-599-223703
8	42	75.0	19662	15	US-10-084-846A-6
9	41	73.2	94	12	US-10-424-599-285441
10	41	73.2	112	12	US-10-424-599-192615
11	41	73.2	433	12	US-10-112-944-336
12	41	73.2	550	14	US-10-312-187-35
13	41	73.2	747	15	US-10-104-047-2445
14	41	73.2	1471	8	US-08-811-519A-1
15	41	73.2	1474	14	US-10-225-567A-522

16	41	73.2	1474	15	US-10-292-798-914	Sequence 914, App
17	41	73.2	1515	14	US-10-240-154-8	Sequence 8, Appli
18	41	73.2	2796	9	US-09-870-759-114	Sequence 114, App
19	41	73.2	2796	10	US-09-751-708A-114	Sequence 114, App
20	40	71.4	229	12	US-10-425-114-62147	Sequence 62147, A
21	40	71.4	326	9	US-09-815-243-10259	Sequence 10259, A
22	40	71.4	326	9	US-09-815-242-14087	Sequence 14087, A
23	40	71.4	517	12	US-10-425-114-43375	Sequence 43375, A
24	40	71.4	556	14	US-10-156-781-9809	Sequence 9809, Ap
25	39	69.6	71	12	US-10-424-599-178111	Sequence 178111, A
26	39	69.6	146	12	US-10-425-114-49847	Sequence 49847, A
27	39	69.6	164	12	US-10-424-599-174830	Sequence 174830, A
28	39	69.6	183	12	US-10-424-599-268092	Sequence 268092, A
29	39	69.6	198	12	US-10-424-599-239718	Sequence 239718, A
30	39	69.6	217	12	US-10-425-114-45354	Sequence 45354, A
31	39	69.6	284	12	US-10-653-595-2398	Sequence 2398, App
32	39	69.6	278	12	US-10-425-114-66403	Sequence 66403, A
33	39	69.6	335	9	US-09-821-687-11	Sequence 11, Appl
34	39	69.6	423	12	US-10-425-114-59390	Sequence 59390, A
35	39	69.6	470	12	US-10-283-132A-62669	Sequence 62669, A
36	39	69.6	618	14	US-10-408-209-5	Sequence 5, Appli
37	39	69.6	633	9	US-09-821-687-10	Sequence 10, Appl
38	39	69.6	633	10	US-09-374-046A-128	Sequence 128, App
39	39	69.6	633	12	US-10-616-263-128	Sequence 128, App
40	38	67.9	66	14	US-10-029-386-39711	Sequence 29711, A
41	38	67.9	86	12	US-10-424-599-264575	Sequence 264575, A
42	38	67.9	110	12	US-10-424-599-285656	Sequence 285656, A
43	38	67.9	111	12	US-10-424-599-244473	Sequence 244473, A
44	38	67.9	137	12	US-10-424-599-150525	Sequence 150525, A
45	38	67.9	145	12	US-10-424-599-246033	Sequence 246033, A

## ALIGNMENTS

## RESULT 1

US-09-864-761-43644  
; Sequence 43644, Application US/09864761  
; Patent NO. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30

```
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43644
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006518.17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77
; OTHER INFORMATION: EST HUMAN HIT: BF088785.1, EVALUATE 1.00e-01
; OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUATE 7.40e-02
US-09-864-761-43644

Query Match 100.0%; Score 56; DB 9; Length 132;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHPRPGR 9
Db 82 GHPRPGR 90

RESULT 2
US-10-157-031-80
; Sequence 80, Application US/10157031
; Publication No. US20030108890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krutovskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-157-031-80

Query Match 100.0%; Score 56; DB 14; Length 166;
Best Local Similarity 100.0%; Pred. No. 2.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHPRPGR 9
Db 116 GHPRPGR 124

RESULT 3
US-10-084-846A-7
; Sequence 7, Application US/10084846A

; PRIOR APPLICATION NUMBER: PCT/US01/00626A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 7
; LENGTH: 19652
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 2.
; OTHER INFORMATION: Start codon: gat, Start position: nucleotide 2.
US-10-084-846A-7

Query Match 82.1%; Score 46; DB 15; Length 19652;
Best Local Similarity 88.9%; Pred. No. 3.4e+03;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GHPRPGR 9
Db 2258 GTPRPGR 2266

RESULT 4
US-10-425-114-69167
; Sequence 69167, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 69167
; LENGTH: 238
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLM017307C10_FLI.pep
US-10-425-114-69167

Query Match 80.4%; Score 45; DB 12; Length 238;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 HPRPGR 8
Db 155 HPRPGR 161

RESULT 5
US-10-425-114-43478
; Sequence 43478, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; PENDING FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 43478
; LENGTH: 277
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700423863_FLI.pap
US-10-425-114-43478

Query Match      80.4%; Score 45; DB 12; Length 277;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 HPRPPRG 8
DB      86 HPRPPRG 92

RESULT 6
US-09-864-761-45389
; Sequence 45389; Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
```

```
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/508,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 45389
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP001427.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.77
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.75
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.55
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.71
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.61
; OTHER INFORMATION: SWISSPROT HIT: O15016, EVALUATE 3.70e+00
; OTHER INFORMATION: EST_HUMAN HIT: BF525926.1, EVALUATE 2.40e+00
US-09-864-761-45389

Query Match      75.0%; Score 42; DB 9; Length 40;
Best Local Similarity 87.5%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 GHRPPRG 8
DB      31 GHRPPRG 38

RESULT 7
US-10-424-599-223703
; Sequence 223703; Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 223703
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_44033C.1.pap
US-10-424-599-223703

Query Match      75.0%; Score 42; DB 12; Length 72;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3 PRPPRG 9
DB      27 PRPPRG 33

RESULT 8
US-10-084-846A-6
; Sequence 6; Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
```

; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
; FILE REFERENCE: 1974-005  
; CURRENT APPLICATION NUMBER: US/10/084,846A  
; CURRENT FILING DATE: 2003-02-25  
; PRIOR APPLICATION NUMBER: PCT/EP01/09815  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: DE 101 09 166.4  
; PRIOR FILING DATE: 2001-02-25  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: Patentin Ver. 3.2  
; SEQ ID NO 6  
; LENGTH: 19662  
; TYPE: PRT  
; ORGANISM: Streptomyces viridochromogenes  
; FEATURE:  
; OTHER INFORMATION: Protein 1: amino acid sequence encoded by coding strand 2.  
; OTHER INFORMATION: Start codon: gga, Start position: nucleotide 1.  
US-10-084-846A-6

Query Match 75.0%; Score 42; DB 15; Length 19662;  
Best Local Similarity 77.8%; Pred. No. 1.2e+04;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GHPRPPRGR 9  
| | | | |  
Db 13062 GRPRPPRSR 13070

## RESULT 9

US-10-424-599-285441  
; Sequence 285441, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 285441  
; LENGTH: 94  
; TYPE: PRT  
; ORGANISM: Glycine max

; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_99782C.1.pap  
US-10-424-599-285441

Query Match 73.2%; Score 41; DB 12; Length 94;  
Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GHPRPPRGR 9  
| | | | |  
Db 39 GIPQPPRGR 47

## RESULT 10

US-10-424-599-192615  
; Sequence 192615, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 192615  
; LENGTH: 112  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_15954C.1.pap  
US-10-424-599-192615

Query Match 73.2%; Score 41; DB 12; Length 112;  
Best Local Similarity 77.8%; Pred. No. 1.8e+02;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GHPRPPRGR 9  
| | | | |  
Db 103 GHPTPRGR 111

## RESULT 11

US-10-112-944-336  
; Sequence 336, Application US/10112944  
; Publication No. US20040048249A1  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Y. Tom  
; APPLICANT: Yang, Yonghong  
; APPLICANT: Weng, Gezhi  
; APPLICANT: Zhang, Jie  
; APPLICANT: Ren, Feiyan  
; APPLICANT: Xue, Aidong J.

; APPLICANT: Wang, Jian-Rui  
; APPLICANT: Wehrman, Tom  
; APPLICANT: Ghosh, Malabika  
; APPLICANT: Wang, Dunrui  
; APPLICANT: Zhao, Qing A.  
; APPLICANT: Wang, Zhiwei  
; TITLE OF INVENTION: No. US20040048249A1el Nucleic Acids and  
; FILE REFERENCE: 805A  
; CURRENT APPLICATION NUMBER: US/10/112,944  
; CURRENT FILING DATE: 2002-03-28  
; PRIOR APPLICATION NUMBER: US 09/488,725  
; PRIOR FILING DATE: 2000-01-21  
; PRIOR APPLICATION NUMBER: US 09/491,404  
; PRIOR FILING DATE: 2000-01-25  
; PRIOR APPLICATION NUMBER: US 09/496,914  
; PRIOR FILING DATE: 2000-02-03  
; PRIOR APPLICATION NUMBER: US 09/515,126  
; PRIOR FILING DATE: 2000-02-28  
; PRIOR APPLICATION NUMBER: US 09/519,705  
; PRIOR FILING DATE: 2000-03-07  
; PRIOR APPLICATION NUMBER: US 09/540,217  
; PRIOR FILING DATE: 2000-03-31  
; PRIOR APPLICATION NUMBER: US 09/552,929  
; PRIOR FILING DATE: 2000-04-18  
; PRIOR APPLICATION NUMBER: US 09/577,408  
; PRIOR FILING DATE: 2000-05-18  
; NUMBER OF SEQ ID NOS: 924  
; SOFTWARE: pt\_FL\_genes Version 5.0  
; SEQ ID NO 336  
; LENGTH: 433  
; TYPE: PRT  
; ORGANISM: Homo sapiens

US-10-112-944-336

Query Match 73.2%; Score 41; DB 12; Length 433;  
Best Local Similarity 85.7%; Pred. No. 5.9e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GHPRPPR 7  
| | | | |  
Db 68 GHPRPPQ 74

```

RESULT 12
US-10-312-187-35
; Sequence 35, Application US/10312187
; Publication No. US20030162291A1
; GENERAL INFORMATION:
; APPLICANT: HEKIMI, Siegfried
; APPLICANT: BENARD, Claire
; APPLICANT: MCGRIGHT, Brent
; APPLICANT: LAKOWSKI, Bernard
; APPLICANT: HAN, Dong
; APPLICANT: LABBE, Jean-Claude
; TITLE OF INVENTION: CLK-2, CEX-7 AND COQ-4 GENES, AND USES THEREOF
; FILE REFERENCE: 11202-008-999
; CURRENT APPLICATION NUMBER: US/10/312,187
; CURRENT FILING DATE: 2002-12-20
; PRIOR APPLICATION NUMBER: US 60/254,932
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/213,174
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 52
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 550
; TYPE: PRT
; ORGANISM: unknown
; FEATURE:
; OTHER INFORMATION: XE7, Homo sapiens cex-7 protein
US-10-312-187-35

```

```

Query Match      73.2%; Score 41; DB 14; Length 550;
Best Local Similarity 77.8%; Pred. No. 7.3e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 GHPRPGR 9
DB 497 GEPGPPRGR 505

```

```

RESULT 13
US-10-104-047-2445
; Sequence 2445, Application US/10104047
; Publication No. US20030236392A1
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: NO. US20030236392A1 full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; CURRENT FILING DATE: 2002-03-25
; PRIOR FILING DATE:
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2445
; LENGTH: 747
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2445

```

```

Query Match      73.2%; Score 41; DB 15; Length 747;
Best Local Similarity 85.7%; Pred. No. 9.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 GHPRPGR 7
DB 307 GHPRPQ 313

```

```

RESULT 14
US-08-811-519A-1
; Sequence 1, Application US/08811519A
; Publication No. US20030143665A1
; GENERAL INFORMATION:

```

```

; APPLICANT: Petrenko, Alexandre
; TITLE OF INVENTION: CALCIUM INDEPENDENT RECEPTOR OF
; NUMBER OF INVENTIONS: ALPHA-LATROTOXIN, CHARACTERIZATION AND USES THEREOF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David A. Jackson, Esq.
; STREET: 411 Hackensack Ave, Continental Plaza, 4th
; STREET: Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,519A
; FILING DATE: 4-MAR-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 1049-1-007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1471 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
US-08-811-519A-1

```

```

Query Match      73.2%; Score 41; DB 8; Length 1471;
Best Local Similarity 77.8%; Pred. No. 1.7e+03;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 GHPRPGR 9
DB 1263 GGEPPRGR 1271

```

```

RESULT 15
US-10-225-567A-522
; Sequence 522, Application US/10225567A
; Publication No. US20030113798A1
; GENERAL INFORMATION:
; APPLICANT: LifeSpan Biosciences
; APPLICANT: Brown, Joseph P.
; APPLICANT: Burner, Glenna C.
; APPLICANT: Roush, Christine L.
; TITLE OF INVENTION: ANTIGENIC PEPTIDES AND ANTIBODIES FOR G PROTEIN-COUPLED RECEPTOR
; FILE REFERENCE: 1920-4-4
; CURRENT APPLICATION NUMBER: US/10/225,567A
; CURRENT FILING DATE: 2001-12-19
; PRIOR APPLICATION NUMBER: 60/257,144
; PRIOR FILING DATE: 2000-12-19
; NUMBER OF SEQ ID NOS: 2292
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 522
; LENGTH: 1474
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-225-567A-522

```

```

Query Match      73.2%; Score 41; DB 14; Length 1474;
Best Local Similarity 77.8%; Pred. No. 1.7e+03;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

Qy 1 GHPPPPRGR 9  
Db 1264 GGPPPPRGR 1272

Search completed: April 6, 2004, 17:06:06  
Job time : 42.3271 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:52:34 ; Search time 8.97196 Seconds  
(without alignments)  
85.771 Million cell updates/sec

Title: US-10-009-709-9

Perfect score: 50

Sequence: 1 HRPFRGR 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	166	1 PIRUSC	salivary proline-r
2	50	100.0	166	2 B23372	salivary proline-r
3	50	100.0	171	2 A27307	proline-rich phosph
4	44	88.0	561	2 E70610	hypothetical prote
5	41	82.0	2796	2 JC4743	fatty-acid synthas
6	40	80.0	129	2 F71241	hypothetical prote
7	40	80.0	691	2 A54741	erythrocyte membra
8	39	78.0	200	2 S53609	hypothetical prote
9	39	78.0	346	2 I48185	gene alx3 protein
10	39	78.0	470	2 C70641	hypothetical prote
11	39	78.0	3649	1 S18268	delta-(L-alpha-ami
12	38	76.0	450	2 AB3238	nitrilotriacetate
13	37	74.0	92	2 I70113	C-arrestin - rat (
14	37	74.0	107	2 A72701	hypothetical prote
15	37	74.0	123	2 D72579	hypothetical prote
16	37	74.0	137	2 A88637	protein W09G12.9 [
17	37	74.0	151	2 H82546	hypothetical prote
18	37	74.0	237	2 D40595	hypothetical prote
19	37	74.0	256	1 ASLJH2	nef protein - huma
20	37	74.0	340	2 A24026	erythromycin resis
21	37	74.0	401	2 T32737	hypothetical prote
22	37	74.0	429	1 FOLJGH	probable gag poly
23	37	74.0	429	1 FOLJGH	gag polyprotein -
24	37	74.0	429	2 S06073	gag polyprotein -
25	37	74.0	438	2 B72654	probable histidyl-
26	37	74.0	513	2 B87484	anthranilate synth
27	37	74.0	1385	2 H88569	protein K03H1.5 [I
28	37	74.0	1409	2 S41028	hypothetical prote
29	37	74.0	1466	2 T17138	CLiAA protein - ra

#### ALIGNMENTS

##### RESULT 1

##### PIRUSC

salivary proline-rich phosphoprotein precursor PRH2 [validated] - human

N;Alternate names: salivary acidic proline-rich protein PRH2

C;Contains: peptide P-C (basic proline-rich peptide IB-8b); proline-rich phosphoprotein

C;Species: Homo sapiens (man)

C;Date: 31-Mar-1981 #sequence revision 12-Apr-1996 #text change 08-Dec-2000

C;Accession: A25372; A19803; B57868; A92277; A92254; A94125; A91954; S02564; S02563; JP

R;Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.

J. Biol. Chem. 260, 11123-11130, 1985

A;Title: Differential RNA splicing and post-translational cleavages in the human saliva

A;Reference number: A92492; MUID:85289325; PMID:2993301

A;Accession: A25372

A;Molecule type: mRNA

A;Residues: 1-166 <MA>

A;Cross-references: GB:K03202; NID:G130481; PIDN:AAA60183.1; PID:G190482

R;Schlesinger, D.H.; Hay, D.I.

Int. J. Pept. Protein Res. 17, 34-41, 1981

A;Title: Primary structure of the active tryptic fragments of human and monkey salivary

A;Reference number: A91757; MUID:81191179; PMID:7238490

A;Accession: A19803

A;Molecule type: protein

A;Residues: 17-46 <SCH>

R;Kim, H.S.; Maeda, N.

J. Biol. Chem. 261, 6712-6718, 1986

A;Title: Structures of two HaeIII-type genes in the human salivary proline-rich protein

A;Reference number: A57868; MUID:86196106; PMID:3009472

A;Accession: B57868

A;Molecule type: DNA

A;Residues: 1-166 <KIM>

A;Cross-references: GB:M13058; NID:G190513; PIDN:AAA98808.1; PID:G190514

R;Wong, R.S.C.; Bennick, A.

J. Biol. Chem. 255, 5943-5948, 1980

A;Title: The primary structure of a salivary calcium-binding proline-rich phosphoprotei

A;Reference number: A92277; MUID:80204368; PMID:7380845

A;Contents: protein C

A;Accession: A92277

A;Molecule type: protein

A;Residues: 17-19, 'N', '21-166 <WON>

A;Note: the amino-terminal 46 residues are involved with inhibiting hydroxyapatite form

R;Wong, R.S.C.; Hofmann, T.; Bennick, A.

J. Biol. Chem. 254, 4800-4808, 1979

A;Title: The complete primary structure of a proline-rich phosphoprotein from human sal

A;Reference number: A92254; MUID:79173237; PMID:438215

A;Contents: protein A

A;Accession: A92254

A;Molecule type: protein

A;Residues: 17-19, 'N', '21-122 <WO2>

R;Schlesinger, D.H.; Hay, D.I.

In Peptides: Structure and Biological Function (Proc. Sixth Amer. Peptide Symp.), Gross

A;Title: Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent

A;Reference number: A94425

latrophilin-1, bra  
CLiBA protein - ra  
latrophilin-1, bra  
CLiAB protein - ra  
CLiBB protein - ra  
cysteine proteinas  
oryzacystatin II -  
hypothetical prote  
hypothetical prote  
protein F3F9.20 [I  
cholecystokinin B  
conserved hypothes  
gastatin/cholecysto  
alpha-2B-adrenargi  
gastrin receptor -  
cholecystokinin B

30 37 74.0 1467 2 T18411  
31 37 74.0 1471 2 T17149  
32 37 74.0 1472 2 T18413  
33 37 74.0 1510 2 T17145  
34 37 74.0 1515 2 T17156  
35 36 72.0 106 2 S13027  
36 36 72.0 107 2 A38375  
37 36 72.0 167 2 A2630  
38 36 72.0 264 2 S75053  
39 36 72.0 302 2 H96811  
40 36 72.0 381 2 S48049  
41 36 72.0 444 2 AC3162  
42 36 72.0 447 2 A47430  
43 36 72.0 448 2 I51883  
44 36 72.0 450 2 J01614  
45 36 72.0 452 2 A46195

A;Accession: A94425  
 A;Molecule type: protein  
 A;Residues: 17-122 <SC2>  
 A;Note: the authors call this protein PRP-4  
 J;Isemura, S.; Saitoh, E.; Sanada, K.  
 J. Biol. Chem. 260, 1071-1077, 1985  
 A;Title: The amino acid sequence of a salivary proline-rich peptide, P-C, and its relationship to the human saliva  
 A;Reference number: A91954; MUID:80227634; PMID:7390979  
 A;Contents: peptide P-C  
 A;Accession: A91954  
 A;Molecule type: protein  
 A;Residues: 123-166 <RSE>  
 R;Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck  
 Biochem. J. 255, 15-21, 1988  
 A;Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-1 to PRP-6)  
 A;Reference number: S02562; MUID:89061650; PMID:3196309  
 A;Accession: S02564  
 A;Molecule type: protein  
 A;Residues: 17-166 <HAY>  
 A;Accession: S02563  
 A;Molecule type: protein  
 A;Residues: 47-71 <HA2>  
 R;Schlesinger, D.H.; Hay, D.I.  
 Int. J. Pept. Protein Res. 27, 373-379, 1986  
 A;Title: Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of the human salivary acidic proline-rich phosphoprotein, PRP-1  
 A;Reference number: JP0106; MUID:86222916; PMID:3710693  
 A;Accession: JP0106  
 A;Molecule type: protein  
 A;Residues: 17-161 'Q' 163-166 <SC3>  
 A;Experimental source: parotid gland  
 R;Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A;Title: Basic proline-rich proteins from human parotid saliva: relationships of the covalent structure and possible origin of the non-glycosylated basic proline-rich protein  
 A;Reference number: A38355; MUID:91190884; PMID:1849422  
 A;Accession: G38355  
 A;Status: preliminary  
 A;Molecule type: protein  
 A;Residues: 123-166 <KAU>  
 R;Robinson, R.; Kaufman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A;Title: Primary structure and possible origin of the non-glycosylated basic proline-rich protein  
 A;Reference number: S06153; MUID:90088384; PMID:2688632  
 A;Accession: S06153  
 A;Molecule type: protein  
 A;Residues: 123-166 <ROB>  
 R;Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A;Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich protein  
 A;Reference number: A27307; MUID:88074309; PMID:3687941  
 A;Contents: allele Pa  
 A;Accession: B27307  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: DNA  
 A;Residues: 17-41 'L' 43-118 'C' 120-166 <AZE>  
 A;Cross-references: EMBL:K03203  
 C;Genetics:  
 A;Gene: GDB:PRH1  
 A;Cross-references: GDB:119516; OMIM:168790  
 A;Map position: 12p13.2-12p13.2  
 A;Introns: 22/1, 34/1  
 C;Superfamily: proline-rich protein  
 C;Keywords: proline-rich protein; pyroglutamic acid; saliva  
 F;1-16/Domain: signal sequence #status predicted <SIG>  
 F;17-166/Product: protein C #status experimental <PRC>  
 F;17-122/Product: protein A #status experimental <PRA>  
 F;17-46/Region: apatitic mineral binding  
 F;47-71/Product: PRP-3 #status experimental <PRP3>  
 F;123-166/Product: peptide P-C #status experimental <PPC>  
 F;17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimental  
 F;24,38/Binding site: phosphate (Ser) (covalent) #status experimental  
 Query Match 100.0%; Score 50; DB 1; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 0.45; Mismatches 0; Indels 0; Gaps 0;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HRPFRGR 8  
 DB 117 HRPFRGR 124  
 RESULT 2  
 B25372  
 proline-rich phosphoprotein (gene PRH1, Db allele) - human  
 N;Alternate names: salivary acidic proline-rich protein  
 C;Species: Homo sapiens (man)

salivary proline-rich phosphoprotein precursor PRH1 (allele PIR) - human  
 C;Species: Homo sapiens (man)  
 C;Date: 29-Aug-1987 #sequence, revision 29-Aug-1987 #text change 20-Aug-1999  
 A;Accession: B25372; A57868; S02562; G38355; S06153; B27307  
 R;Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
 J. Biol. Chem. 260, 11123-11130, 1985  
 A;Title: Differential RNA splicing and post-translational cleavages in the human saliva  
 A;Reference number: A92492; MUID:85289325; PMID:2993301  
 A;Accession: B25372  
 A;Molecule type: mRNA  
 A;Residues: 1-166 <MAE>  
 A;Cross-references: GB:K03203; NID:G190483; PIDN:AAA60184.1; PID:G190484  
 R;Kim, H.S.; Maeda, N.  
 J. Biol. Chem. 261, 6712-6718, 1986  
 A;Title: Structures of two HaeIII-type genes in the human salivary proline-rich protein  
 A;Reference number: A57868; MUID:86196106; PMID:3009472  
 A;Accession: A57868  
 A;Molecule type: DNA  
 A;Residues: 1-166 <KIM>  
 A;Cross-references: GB:M13057; NID:G190511; PIDN:AAA98807.1; PID:G190512  
 R;Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck  
 Biochem. J. 255, 15-21, 1988  
 A;Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-1 to PRP-6)  
 A;Reference number: S02562; MUID:89061650; PMID:3196309  
 A;Accession: S02562  
 A;Molecule type: protein  
 A;Residues: 47-71 <HAY>  
 R;Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A;Title: Basic proline-rich proteins from human parotid saliva: relationships of the covalent structure and possible origin of the non-glycosylated basic proline-rich protein  
 A;Reference number: A38355; MUID:91190884; PMID:1849422  
 A;Accession: G38355  
 A;Molecule type: protein  
 A;Residues: 123-166 <KAU>  
 R;Robinson, R.; Kaufman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A;Title: Primary structure and possible origin of the non-glycosylated basic proline-rich protein  
 A;Reference number: S06153; MUID:90088384; PMID:2688632  
 A;Accession: S06153  
 A;Molecule type: protein  
 A;Residues: 123-166 <ROB>  
 R;Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A;Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich protein  
 A;Reference number: A27307; MUID:88074309; PMID:3687941  
 A;Contents: allele Pa  
 A;Accession: B27307  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: DNA  
 A;Residues: 17-41 'L' 43-118 'C' 120-166 <AZE>  
 A;Cross-references: EMBL:K03203  
 C;Genetics:  
 A;Gene: GDB:PRH1  
 A;Cross-references: GDB:119515; OMIM:168730  
 A;Map position: 12p13.2-12p13.2  
 A;Introns: 22/1, 34/1  
 C;Superfamily: proline-rich protein  
 C;Keywords: phosphoprotein; saliva; tandem repeat  
 Query Match 100.0%; Score 50; DB 2; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 0.45; Mismatches 0; Indels 0; Gaps 0;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HRPFRGR 8  
 DB 117 HRPFRGR 124  
 RESULT 3  
 A27307  
 proline-rich phosphoprotein (gene PRH1, Db allele) - human  
 N;Alternate names: salivary acidic proline-rich protein  
 C;Species: Homo sapiens (man)

C;Date: 30-Jun-1998 #sequence\_revision 30-Jun-1998 #text\_change 29-Aug-1997  
 C;Accession: A27307  
 C;Species: Mycobacterium tuberculosis  
 A;Title: Alleles at the PHL locus coding for the human salivary-acidic proline-rich protein  
 A;Reference number: A27307; MUID:88074309; PMID:3687941  
 A;Accession: A27307  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: DNA  
 A;Residues: 1-171 <AZE>  
 A;Cross-references: EMBL:K03203  
 C;Genetics:  
 A;Gene: GDB:PRH1  
 A;Cross-references: GDB:113515; OMIM:168730  
 A;Map position: 12p13.2-12p13.2  
 C;Superfamily: proline-rich protein  
 C;Keywords: phosphoprotein

Query Match 100.0%; Score 50; DB 2; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 0.46;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HPRPPRGR 8  
 |||||  
 Db 122 HPRPPRGR 129

RESULT 4  
 E70610  
 hypothetical protein Rv1215c - Mycobacterium tuberculosis (strain H37RV)  
 C;Species: Mycobacterium tuberculosis  
 C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C;Accession: E70610  
 R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
 A;Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A;Reference number: A70500; MUID:98295987; PMID:9634230  
 A;Accession: E70610  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-561 <COL>  
 A;Cross-references: GB:293777; GB:AL123456; NID:g3261726; PIDN:CAB07817.1; PID:e311160;  
 A;Experimental source: strain H37RV  
 C;Genetics:  
 A;Gene: Rv1215c

Query Match 88.0%; Score 44; DB 2; Length 561;  
 Best Local Similarity 87.5%; Pred. No. 13;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 HPRPPRGR 8  
 :|||||  
 Db 534 YPRPPRGR 541

RESULT 5  
 IC4743  
 fatty-acid synthase (EC 2.3.1.85) - Mycobacterium bovis  
 C;Species: Mycobacterium bovis  
 C;Date: 10-May-1996 #sequence\_revision 16-Aug-1996 #text\_change 05-May-2000  
 C;Accession: JC4743  
 R;Fernandes, N.D.; Kolattukudy, P.E.  
 Gene 170, 95-99, 1996  
 A;Title: Cloning, sequencing and characterization of a fatty acid synthase-encoding gene  
 A;Reference number: JC4743; MUID:96200863; PMID:8621098  
 A;Accession: JC4743  
 A;Molecule type: DNA  
 A;Residues: 1-2796 <FER>  
 A;Cross-references: GB:U36763; NID:g1036834; PIDN:AB03809.1; PID:g1036835  
 A;Note: the source is designated as Mycobacterium tuberculosis var. bovis BCG

C;Comment: This enzyme catalyzes both de novo synthesis and chain elongation of fatty acids;  
 C;Genetics:  
 A;Gene:fas  
 C;Superfamily: Mycobacterium tuberculosis fatty-acid synthase  
 C;Keywords: acyltransferase; coenzyme A; phosphopantetheine; phosphoprotein  
 F;2189-2193/Region: nucleotide binding #status predicted  
 F;57/Active site: Ser #status predicted  
 F;1693/Binding site: phosphopantetheine (Ser) (covalent) #status predicted  
 F;2598/Active site: Cys #status predicted

Query Match 82.0%; Score 41; DB 2; Length 2796;  
 Best Local Similarity 87.5%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HPRPGR 8  
 |||||  
 Db 422 HPRPGR 429

RESULT 6  
 F71241  
 hypothetical protein PH0192 - Pyrococcus horikoshii  
 C;Species: Pyrococcus horikoshii  
 C;Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 15-Sep-2000  
 C;Accession: F71241  
 R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Seki, M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguni, DNA Res. 5, 55-76, 1998  
 A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
 A;Reference number: A71000; MUID:98344137; PMID:9679194  
 A;Accession: F71241  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-129 <KAW>  
 A;Cross-references: GB:AP000001; NID:g3236128; PIDN:BAA29261.1; PID:g3256578  
 A;Experimental source: strain ON3  
 A;Note: this accession replaces an interim accession for a sequence replaced by GenBank  
 C;Genetics:  
 A;Gene: PH0192  
 C;Superfamily: Pyrococcus horikoshii hypothetical protein PH0192

Query Match 80.0%; Score 40; DB 2; Length 129;  
 Best Local Similarity 85.7%; Pred. No. 13;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HPRPPRGR 7  
 |||||  
 Db 83 HPRPPRGR 89

RESULT 7  
 AS4741  
 erythrocyte membrane band 4.2 protein - mouse  
 C;Species: Mus musculus (house mouse)  
 C;Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 05-Nov-1999  
 C;Accession: AS4741; I48901  
 R;Korsgren, C.; Cohen, C.M.  
 Genomics 21, 478-485, 1994  
 A;Title: cDNA sequence, gene sequence, and properties of murine pallidin (band 4.2), the  
 A;Reference number: AS4741; MUID:95048323; PMID:7959722  
 A;Accession: AS4741  
 A;Status: preliminary  
 A;Molecule type: mRNA  
 A;Residues: 1-691 <KOR>  
 A;Cross-references: GB:U04055  
 R;Note: authors translated the codon TAC for residue 129 as Ile, and GGT for residue 35  
 R;Rybicki, A.C.; Schwartz, R.S.; Qiu, J.J.; Gilman, J.G.  
 Mamm. Genome 5, 438-445, 1994  
 A;Title: Molecular cloning of mouse erythrocyte protein 4.2: a membrane protein with st  
 A;Reference number: I48901; MUID:95003352; PMID:7919657  
 A;Accession: I48901  
 A;Status: preliminary; translated from GB/EMBL/DBJ  
 A;Molecule type: mRNA

A;Residues: 1-352, 'A', 354-620, 'S', 622-691 <RES>  
A;Cross-references: EMBL:U03487; NID:G424119; PIDN:AAA62275.1; PID:G424120  
C;Superfamily: protein-glutamine gamma-glutamyltransferase  
C;Keywords: blocked amino end; lipoprotein; myristylation  
F;2/Modified site: myristylated amino end (Gly) (in mature form) #status predicted

Query Match 80.0%; Score 40; DB 2; Length 691;  
Best Local Similarity 85.7%; Pred. No. 66;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HPRPPRG 7  
| | | | |  
Db 350 HPRPPNG 356

RESULT 8  
S55609  
hypothetical protein 13 - equine herpesvirus 2  
C;Species: equine herpesvirus 2  
C;Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 08-Oct-1999  
C;Accession: S55609  
E;Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.  
J. Mol. Biol. 249, 520-528, 1995  
A;Title: The DNA sequence of equine herpesvirus 2.  
A;Reference number: S55594; MUID:95302501; PMID:7783207  
A;Accession: S55609  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-200 <TEL>  
A;Cross-references: GB:U20824; NID:G695172; PIDN:AAC13802.1; PID:G695187  
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

Query Match 78.0%; Score 39; DB 2; Length 200;  
Best Local Similarity 85.7%; Pred. No. 29;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 PRPPRGR 8  
| | | | |  
Db 13 PKPPRGR 19

RESULT 9  
I48185  
Gene alx3 protein - golden hamster  
C;Species: Mesocricetus auratus (golden hamster)  
C;Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 24-Sep-1999  
C;Accession: I48185  
R;Rudnick, A.; Ling, T.Y.; Odagiri, H.; Rutter, W.J.; German, M.S.  
Proc. Natl. Acad. Sci. U.S.A. 91, 12203-12207, 1994  
A;Title: Pancreatic beta cells express a diverse set of homeobox genes.  
A;Reference number: I48185; MUID:95083670; PMID:7991607  
A;Accession: I48185  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: mRNA  
A;Residues: 1-346 <RES>  
A;Cross-references: EMBL:X81403; NID:G587455; PIDN:CAA57161.1; PID:G587456  
C;Genetics:  
A;Superfamily: unassigned homeobox proteins; homeobox homology  
C;Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F;157-213/Domain: homeobox homology <HOX>

Query Match 78.0%; Score 39; DB 2; Length 346;  
Best Local Similarity 85.7%; Pred. No. 49;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HPRPPRG 7  
| | | | |  
Db 42 HPAPPRG 48

RESULT 10  
C70641  
hypothetical protein RV0696 - Mycobacterium tuberculosis (strain H37RV)  
C;Species: Mycobacterium tuberculosis  
C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
C;Accession: C70641  
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, R.; Cole, S.T.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Connor, R.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998  
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A;Reference number: A70500; MUID:98295987; PMID:9634230  
A;Accession: C70641  
A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
A;Molecule type: DNA  
A;Residues: 1-470 <COL>  
A;Cross-references: GB:Z84395; GB:AL123456; NID:G3261698; PIDN:CAB06459.1; PID:e293114;  
A;Experimental source: strain H37RV  
C;Genetics:  
A;Gene: RV0696

Query Match 78.0%; Score 39; DB 2; Length 470;  
Best Local Similarity 85.7%; Pred. No. 66;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HPRPPRG 7  
| | | | |  
Db 72 HPRPPSG 78

RESULT 11  
S18268  
delta- (L-alpha-aminoacyl)-L-cysteiny-D-valine synthetase - Streptomyces lactamurans  
C;Species: Streptomyces lactamurans  
C;Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 03-Nov-2000  
C;Accession: S18268; S15283; B38171  
R;Martin, J.F.  
submitted to the EMBL Data Library, January 1991

A;Reference number: S18268  
A;Accession: S18268  
A;Molecule type: DNA  
A;Residues: 1-3649 <MAR>  
A;Cross-references: EMBL:X57310; NID:G45005; PIDN:CAA40561.1; PID:G45006  
R;Coque, J.J.R.; Martin, J.F.; Calzada, J.G.; Liras, P.  
Mol. Microbiol. 5, 1125-1133, 1991  
A;Title: The cephamycin biosynthetic genes pcbAB, encoding a large multidomain peptide  
genes in Acromonium chrysogenum and Penicillium chrysogenum.  
A;Reference number: S15283; MUID:92065808; PMID:1956290  
A;Accession: S15283  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 224-940, 1319-2010, 2373-3307 <COQ>  
A;Cross-references: EMBL:X57310  
A;Note: the source is designated as Nocardia lactamurans  
R;Coque, J.J.R.; Liras, P.; Laiz, L.; Martin, J.F.  
J. Bacteriol. 173, 6258-6264, 1991  
A;Title: A gene encoding lysine 6-aminotransferase, which forms the beta-lactam precursors  
A;Reference number: A38171; MUID:92011390; PMID:1917857  
A;Accession: B38171  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-23 <CO2>  
A;Cross-references: GB:S57006  
C;Genetics:  
A;Gene: pcbAB  
C;Superfamily: alpha-aminoacyl-cysteiny-L-valine synthetase; acetate-CoA ligase homolc  
C;Keywords: carrier protein; cephamycin biosynthesis; phosphopantetheine; phosphoprotei  
F;298-758/Domain: acetate-CoA ligase homology <ACli1>  
F;786-856/Domain: acyl carrier protein homology <ACP1>  
F;1392-1844/Domain: acetate-CoA ligase homology <ACli2>  
F;1862-1932/Domain: acyl carrier protein homology <ACP2>  
F;2446-2895/Domain: acetate-CoA ligase homology <ACli3>  
F;2912-2980/Domain: acyl carrier protein homology <ACP3>  
F;820,1896,2944/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

```

Query Match      78.0%; Score 39; DB 1; Length 3649;
Best Local Similarity 100.0%; Pred. No. 4; 7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      1 HRPFRG 6
      |||||
Db      2153 HRPFRG 2158

RESULT 12
AB3238
nitroloacetate monooxygenase, component A Atu6084 [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AB3238
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.;
  erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
  ; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
  ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AB3238
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-450 <KUR>
A:Cross-references: GB:AE008690; PIDN:AAL46320.1; PID:g17744106; GSPDB:GN00189
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu6084
A:Genome: plasmid
C:Superfamily: nitroloacetate monooxygenase

Query Match      76.0%; Score 38; DB 2; Length 450;
Best Local Similarity 85.7%; Pred. No. 90;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy      2 PRPPGR 8
      |||||
Db      206 PRPPGR 212

RESULT 13
I70113
C-arrestin - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Jun-2000
C:Accession: I70113
R:Craft, C.M.; Whitmore, D.H.; Wiedemann, A.F.
J. Biol. Chem. 269, 4613-4619, 1994
A:Title: Cone arrestin identified by targeting expression of a functional family.
A:Reference number: I55423; MUID:94140898; PMID:8308033
A:Accession: I70113
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-92 <RES>
A:Cross-references: EMBL:U03628; NID:G458204; PIDN:AAAL7552.1; PID:G458205
C:Superfamily: arrestin

Query Match      74.0%; Score 37; DB 2; Length 92;
Best Local Similarity 71.4%; Pred. No. 28;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 HRPFRG 7
      |||||
Db      61 HRPFRG 67

RESULT 14
A72701
hypothetical protein APE1024 - Aeropyrum pernix (strain K1)

```

```

C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: A72701
R:Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Take
  awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
  DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy
  A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: A72701
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-107 <KAW>
A:Cross-references: DBJ:AP000060; NID:G5104188; PIDN:BAAS0009.1; PID:d1043795; PID:G51
  A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1024

Query Match      74.0%; Score 37; DB 2; Length 107;
Best Local Similarity 75.0%; Pred. No. 33;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy      1 HRPFRG 8
      |||||
Db      16 HRPFRG 23

RESULT 15
D72579
hypothetical protein APE1916 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: D72579
R:Kawarabayasi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Take
  awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;
  DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropy
  A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: D72579
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-123 <KAW>
A:Cross-references: DBJ:AP000062; NID:G5105244; PIDN:BAAS0921.1; PID:d1044707; PID:G51
  A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1916

Query Match      74.0%; Score 37; DB 2; Length 123;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy      2 PRPPRG 7
      |||||
Db      31 PRPPRG 36

Search completed: April 6, 2004, 16:16:53
Job time : 8.97196 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd..

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:31:19 ; Search time 5.15888 Seconds  
(without alignments)  
80.746 Million cell updates/sec.

Title: US-10-009-709-9  
Perfect score: 50  
Sequence: 1 HPRPPRGR 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	166	1	PRPC_HUMAN
2	39	78.0	343	1	ALX3_HUMAN
3	39	78.0	343	1	ALX3_MOUSE
4	39	78.0	576	1	Z384_HUMAN
5	39	78.0	579	1	Z384_RAT
6	39	78.0	618	1	MM24_MOUSE
7	39	78.0	3649	1	ACVS_MOUSE
8	38	76.0	260	1	DPML_MOUSE
9	38	76.0	266	1	DPML_CRIGR
10	37	74.0	92	1	ARRC_RAT
11	37	74.0	237	1	YMS5_STRCM
12	37	74.0	256	1	NEF_HV2RO
13	37	74.0	340	1	ERMA_ARTS3
14	37	74.0	429	1	GAG_HTL1A
15	37	74.0	429	1	GAG_HTL1C
16	37	74.0	429	1	GAG_HTL1M
17	37	74.0	438	1	SYH_AERPE
18	37	74.0	687	1	WRK2_AERATH
19	37	74.0	1385	1	YMS5_CABEL
20	37	74.0	1729	1	TABP_HUMAN
21	36	72.0	107	1	CYT2_ORYSA
22	36	72.0	47	1	GASR_HUMAN
23	36	72.0	450	1	GASR_PRANA
24	36	72.0	452	1	GASR_RABIT
25	36	72.0	452	1	GASR_RAT
26	36	72.0	453	1	A2AB_RAT
27	36	72.0	453	1	GASR_CANFA
28	36	72.0	453	1	GASR_MOUSE
29	36	72.0	454	1	GASR_BOVIN
30	36	72.0	455	1	A2AB_MOUSE
31	36	72.0	607	1	MM16_HUMAN
32	36	72.0	607	1	MM16_MOUSE
33	36	72.0	607	1	MM16_RAT

34 36 72.0 703 1 MTAL\_RAT  
35 36 72.0 715 1 MTAL\_HUMAN  
36 36 72.0 715 1 MTAL\_MOUSE  
37 36 72.0 820 1 GLGB\_ORYSA  
38 36 72.0 840 1 SYFB\_STRCO  
39 35 70.0 111 1 YKL1\_YEAST  
40 35 70.0 281 1 YARD\_RHISN  
41 35 70.0 324 1 YSL8\_CHLEN  
42 35 70.0 329 1 Y429\_CHLFR  
43 35 70.0 332 1 Y713\_CHLMU  
44 35 70.0 337 1 TRPD\_HALVO  
45 35 70.0 431 1 ACRO\_RABIT

#### ALIGNMENTS

RESULT 1  
PRPC\_HUMAN STANDARD; PRT; 166 AA.  
AC P02810;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-3) (PRP-2/PRP-4) (PIF-F/PIF-S) (Protein A/protein C) [Contains: Peptide P-C].  
DE Peptide P-C].  
GN PRH1 AND PRH2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ALLELES PRH1-4 AND PRH2-1).  
RX MEDLINE=86196106; PubMed=3009472;  
RA Kim H.-S., Maeda N.;  
RT "Structures of two Haell-type genes in the human salivary proline-rich protein multigene family.";  
RL J. Biol. Chem. 261:6712-6718(1986).  
RN [2]  
RP SEQUENCE FROM N.A. (ALLELES PRH2-1 AND PRH1-PIF).  
RX MEDLINE=85289325; PubMed=293301;  
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
RT "Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system.";  
RL J. Biol. Chem. 260:11123-11130(1985).  
RN [3]  
RP SEQUENCE OF 17-166 (PRP-1 TO PRP-4; PIF-F AND PIF-S).  
RX MEDLINE=8961650; PubMed=3196309;  
RA Hay D.I., Bennick A., Schlesinger D.H., Minaguchi K., Madapallattam G., Schluckebier S.K.;  
RT "The primary structures of six human salivary acidic proline-rich proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-S and PIF-F).";  
RL Biochem. J. 255:15-21(1988).  
RN [4]  
RP SEQUENCE OF 17-166 FROM N.A. (PA ALLELE).  
RX MEDLINE=8074309; PubMed=3687941;  
RA Azen E.A., Kim H.S., Goodman P., Flynn S., Maeda N.;  
RT "Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich proteins Pa, Db, and PIF.";  
RL Am. J. Hum. Genet. 41:1035-1047(1987).  
RN [5]  
RP SEQUENCE OF 17-166 (PRP-2).  
RX MEDLINE=86222916; PubMed=3710693;  
RA Schlesinger D.H., Hay D.I.;  
RT "Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of calcium phosphate crystal growth from human parotid saliva.";  
RL Int. J. Pept. Protein Res. 27:373-379(1986).  
RN [6]  
RP SEQUENCE OF 17-166 (PROTEIN C).  
RX MEDLINE=80204368; PubMed=7380845;  
RA Wong R.S.C., Bennick A.;

Q82599 rattus norv  
Q13330 homo sapien  
Q84480 mus musculu  
Q01401 oryza sativ  
O88054 streptomyce  
P36074 saccharomyc  
P55637 rhizobium s  
Q92834 chlamydia p  
O84436 chlamydia t  
Q9PJW3 chlamydia m  
P52562 halobacteri  
P48038 oryctolagus

RT "The primary structure of a salivary calcium-binding proline-rich  
RT phosphoprotein (protein C), a possible precursor of a related  
RL salivary protein A.";  
RL J. Biol. Chem. 255:5943-5948 (1980).  
RN [7]  
RP SEQUENCE OF 17-46 (PROTEIN C).  
RX MEDLINE=81191179; PubMed=7228490;  
RA Schlesinger D.H., Hay D.I.;  
RT "Primary structure of the active tryptic fragments of human and  
RT monkey salivary anionic proline-rich proteins.";  
RL Int. J. Pept. Protein Res. 17:34-41 (1981).  
RN [8]  
RP SEQUENCE OF 17-122 (PROTEIN A).  
RX MEDLINE=79173237; PubMed=438215;  
RA Wong R.S.C., Hofmann T., Bennick A.;  
RT "The complete primary structure of a proline-rich phosphoprotein from  
RT human saliva.";  
RL J. Biol. Chem. 254:4800-4808 (1979).  
RN [9]  
RP SEQUENCE OF 17-122 (PROTEIN A).  
RX Schlesinger D.H., Hay D.I.;  
RA "Complete primary structure of a proline-rich phosphoprotein (PRP-4),  
RT a potent inhibitor of calcium phosphate precipitation in human parotid  
RT saliva.";  
RL (In) Gross E., Meienhofer J. (eds.);  
RL Peptides: structure and biological function (Proceedings of the 6th  
RL American peptide symposium), pp.133-136, Pierce Chemical Co.,  
RL Rockford IL. (1979).  
RN [10]  
RP SEQUENCE OF 123-166 (PEPTIDE P-C).  
RX MEDLINE=80227634; PubMed=7390979;  
RA Isemura S., Saitoh E., Sanada K.;  
RT "The amino acid sequence of a salivary proline-rich peptide, P-C, and  
RT its relation to a salivary proline-rich phosphoprotein, protein C.";  
RL J. Biochem. 87:1071-1077 (1980).  
RN [11]  
RP VARIANT PRH2-3 LYS-163.  
RA Azen E.A.;  
RT "A frequent mutation in the acidic proline-rich protein gene, PRH2,  
RT causing a Q147K change closely adjacent to the bacterial binding  
RT domain of the cognate salivary PAP (Pr1)' in Afro-Americans.";  
RL Hum. Mutat. 12:72-73 (1998).  
CC -!- FUNCTION: PRP's act as highly potent inhibitors of crystal growth  
CC of calcium phosphates. They provide a protective and reparative  
CC environment for dental enamel which is important for the integrity  
CC of the teeth.  
CC -!- SUBCELLULAR LOCATION: Secreted.  
CC -!- PTM: Proteolytically cleaved; PRP-2, PRP-1, and PIP-S yield PRP-4,  
CC PRP-3 (protein A), and PIP-F, respectively.  
CC -!- POLYMORPHISM: Allele PRH1-4 is also known as PRH1 DA allele;  
CC allele PRH2-1 is also known as PR1 or protein C; allele PRH2-3 is  
CC also known as PR1'.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; X03202; AAA60183.1; -;  
DR EMBL; X03203; AAA60184.1; -;  
DR EMBL; M13057; AAA98807.1; -;  
DR EMBL; M13058; AAA98808.1; -;  
DR Genew; HGNC:9366; PRH1.  
DR Genew; HGNC:9367; PRH2.  
DR MIM; 168730; -;  
DR MIM; 168790; -;  
DR MIM; 168710; -;  
DR GO; GO:0005615; C:extracellular space; TAS.  
DR Repeat; Parotid gland; Phosphorylation; Signal; Polymorphism;  
KW Pyrrolidone carboxylic acid.

FT SIGNAL 1 16  
FT CHAIN 17 166  
FT FT SALIVARY ACIDIC PROLINE-RICH  
FT PHOSPHOPROTEIN 1/2.  
FT CHAIN 17 122  
FT FT SALIVARY ACIDIC PROLINE-RICH  
FT PHOSPHOPROTEIN 3/4.  
FT CHAIN 123 166  
FT DOMAIN 17 46  
FT FT INHIBIT HYDROXYAPATITE FORMATION, BIND  
FT TO HYDROXYAPATITE AND CALCIUM.  
FT MOD RES 17 17  
FT MOD RES 24 24  
FT MOD RES 38 38  
FT VARIANT 20 20  
FT FT D -> N (in allele PRH1-4).  
FT FT /FTId=VAR\_005563.  
FT FT D -> N (in allele PRH2-1).  
FT FT /FTId=VAR\_005564.  
FT FT Q -> K (in allele PRH2-3).  
FT FT /FTId=VAR\_005565.  
FT FT F -> P (IN REF. 10).  
FT CONFLICT 41 41  
SQ SEQUENCE 166 AA; 17017 MW; A7DF62BF94E3C3EF CRC64;  
  
Query Match 100.0%; Score 50; DB 1; Length 166;  
Best Local Similarity 100.0%; Pred. No. 0.19; 0; Indels 0; Gaps 0;  
Matches 8; Conservative 0; Mismatches 0;  
  
QY 1 HRPFRPRGR 8  
DB 117 HRPFRPRGR 124  
|||||||  
|  
  
RESULT 2  
ALX3 HUMAN STANDARD; PRT; 343 AA.  
ID \_ALX3 HUMAN  
AC O95075; O95075;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Homeobox protein aristal-less-like 3 (Proline-rich transcription factor  
DE ALX3).  
DN ALX3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Neuroblastoma;  
RX MEDLINE=21665991; PubMed=11807986;  
RA Wimmer K., Zhu X.-X., Rouillard J.M., Ambros P.F., Lamb B.J.,  
RA Kuick R., Eckart M., Weinhausl A., Fonatsch C., Hanash S.M.;  
RT "Combined restriction landmark genomic scanning and virtual genome  
RT scans identify a novel human homeobox gene, ALX3, that is  
RT hypermethylated in neuroblastoma.";  
RL Genes Chromosomes Cancer 33:285-294 (2002).  
CC -!- FUNCTION: transcriptional regulator with a possible role in  
CC patterning of mesoderm during development (By similarity).  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- SIMILARITY: Belongs to the paired homeobox family.  
CC -!- SIMILARITY: Contains 1 homeobox domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF008203; AAD01418.1; -;  
DR EMBL; AF008202; AAD01417.2; -;  
DR HSSP; P06601; fJUL.  
DR TRANSFAC; T04515; -;  
DR Genew; HGNC:449; ALX3.  
DR MIM; 606014; -;

DR InterPro: IPR001356; Homeobox.  
DR InterPro: IPR007104; Paired\_homeo.  
DR Pfam: PF00046; homeobox; 1.  
DR ProDom: PD000010; Homeobox; 1.  
DR SMART: SM00389; HOX; 1.  
DR PROSITE: PS00027; HOMEBOX 1; 1.  
DR PROSITE: PS00071; HOMEBOX 2; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
KW Transcription regulation.  
FT DNA\_BIND 153 212 HOMEBOX.  
FT CONFLICT 72 72 L -> M (IN REF. 1; AAD01417).  
FT CONFLICT 86 86 F -> L (IN REF. 1; RAD01417).  
SQ SEQUENCE 343 AA; 36904 MW; 17AFC7ECD40B942F CRC64;  
  
Query Match 78.0%; Score 39; DB 1; Length 343;  
Best Local Similarity 85.7%; Pred.No.22;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 HRPFRPG 7  
DB 39 HPAPPRG 45  
  
RESULT 3  
ALX3 MOUSE  
ID ALX3 MOUSE STANDARD; PRT; 343 AA.  
AC 070137;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Homeobox protein aristaless-like 3 (Proline-rich transcription factor  
DE ALX3).  
GN ALX3.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NIH Swiss;  
RX MEDLINE=98340878; PubMed=96761699;  
RA ten Berge D., Brouwer A., el Bahi S., Guenet J.-L., Robert B.,  
RA Weijlink F.;  
RT "Mouse Alx3: an aristaless-like homeobox gene expressed during  
RT embryogenesis in ectomesenchyme and lateral plate mesoderm."  
RL Dev. Biol. 199;11:25(1998).  
CC -!- FUNCTION: Transcriptional regulator with a possible role in  
CC patterning of mesoderm during development.  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- TISSUE SPECIFICITY: Predominantly in neural crest-derived  
CC mesenchyme and in lateral plate mesoderm. Prominent expression in  
CC frontonasal head mesenchyme and in the first and second pharyngeal  
CC arches and some of their derivatives. High expression is also seen  
CC in the tail and in many derivatives of the lateral plate mesoderm  
CC including the limbs, the body wall, and the genital tubercle.  
CC -!- DEVELOPMENTAL STAGE: Expressed in embryos from 8 days of gestation  
CC onward.  
CC -!- SIMILARITY: Belongs to the paired homeobox family.  
CC -!- SIMILARITY: Contains 1 homeobox domain.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC -----  
CC EMBL: U96109; AAC15094.1; --  
CC HSPSP; P06601; 1FJL.  
CC TRANSFAC: T03343; --  
CC MGD; MGI:1277097; Alx3.  
DR InterPro: IPR001356; Homeobox.

DR InterPro: IPR007104; Paired\_homeo.  
DR Pfam: PF00046; homeobox; 1.  
DR ProDom: PD000010; Homeobox; 1.  
DR SMART: SM00389; HOX; 1.  
DR PROSITE: PS00027; HOMEBOX 1; 1.  
DR PROSITE: PS00071; HOMEBOX 2; 1.  
KW Homeobox; DNA-binding; Developmental protein; Nuclear protein;  
KW Transcription regulation.  
FT DNA\_BIND 153 212 HOMEBOX.  
FT CONFLICT 72 72 L -> M (IN REF. 1; AAD01417).  
FT CONFLICT 86 86 F -> L (IN REF. 1; RAD01417).  
SQ SEQUENCE 343 AA; 36950 MW; 87900BF977FAC2D2D CRC64;  
  
Query Match 78.0%; Score 39; DB 1; Length 343;  
Best Local Similarity 85.7%; Pred.No.22;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
  
QY 1 HRPFRPG 7  
DB 39 HPAPPRG 45  
  
RESULT 4  
Z384 HUMAN  
ID Z384 HUMAN STANDARD; PRT; 576 AA.  
AC Q8TF68; O15407; Q8N938;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Zinc finger protein 384 (Nuclear matrix transcription factor 4)  
DE (CAG repeat protein 1).  
GN ZNF384 OR NMP4 OR CAGH1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RA Matsuo M.Y.;  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Brain;  
RA Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,  
RA Wakebe H., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,  
RA Irie R., Otsuki T., Sato H., Wakamatsu A., Iehi S., Yamamoto J.,  
RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,  
RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,  
RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K.,  
RA Masuho Y., Nagai K., Isogai T.;  
RL "NEO human cDNA sequencing project."  
RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE OF 395-576 FROM N.A.  
RC TISSUE=Brain cortex;  
RX MEDLINE=97369492; PubMed=9225980;  
RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.-H., Kidwai A.S.,  
RA Brechel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;  
RT "cDNAs with long CAG trinucleotide repeats from human brain."  
RL Hum. Genet. 100;114-122(1997).  
CC -!- FUNCTION: Transcription factor that binds the consensus DNA  
CC sequence (GC)AAAAA. Seems to bind and regulate the promoters of  
CC MMP1, MMP3, MMP7 and COL1A1 (By similarity).  
CC -!- SUBUNIT: Interacts with Cas (By similarity).  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=2;  
CC Comment=Additional isoforms seem to exist;  
CC Name=1;  
CC IsoID=Q8TF68-1; Sequence=Displayed;  
CC Name=2;  
CC IsoID=Q8TF68-2; Sequence=VSP\_006920;  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.



```

CC -1- SIMILARITY: Contains 8 C2H2-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB070238; BAB85125.1; -
DR EMBL; AK095734; BAC04618.1; -
DR EMBL; U80739; AAB91437.1; -
DR Genew; HGNC:11955; ZNF384.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR ProDom; PD000003; Znf_C2H2; 4.
DR SMART; SM00355; Znf_C2H2; 8.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 8.
KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;
KW DNA-binding; Repeat; Alternative splicing.
FT ZN_FING 228 250 C2H2-TYPE 1.
FT ZN_FING 256 278 C2H2-TYPE 2.
FT ZN_FING 284 306 C2H2-TYPE 3.
FT ZN_FING 317 339 C2H2-TYPE 4.
FT ZN_FING 345 367 C2H2-TYPE 5.
FT ZN_FING 373 397 C2H2-TYPE 6.
FT ZN_FING 403 425 C2H2-TYPE 7.
FT ZN_FING 433 455 C2H2-TYPE 8.
FT DOMAIN 461 521 GLN-RICH.
FT DOMAIN 466 499 ALA-RICH.
FT VARSPLIC 300 360 Missing (in isoform 2).
FT -----
SQ SEQUENCE 576 AA; 63091 MW; 2A52786C3C46D90 CRC64;
Query Match 78.0%; Score 39; DB 1; Length 576;
Best Local Similarity 85.7%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 PRPRGR 8
Db 188 KPPRGR 194
|:|||||
|:|||||
RESULT 5
Z384 RAT STANDARD; PRT; 579 AA.
AC Q9EQJ4; Q9EQJ2; Q9EQJ3; Q9JMJ5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 384 (Nuclear matrix transcription factor 4)
DE (Cas-associated zinc finger protein).
DE ZNF384 OR NMP4 OR CIZ.
GN Rattus norvegicus (Rat).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), SUBCELLULAR LOCATION, DNA-BINDING, AND
RP INTERACTION WITH CAS.
RX MEDLINE=20136045; PubMed=10669742;
RA Nakamoto T., Yamagata T., Sakai R., Ogawa S., Honda H., Ueno H.,
RA Hirano N., Yazaki Y., Hirai H.;
RT "CIZ, a zinc finger protein that interacts with p30Cas and activates
RT the expression of matrix metalloproteinases.";
RL Mol. Cell. Biol. 20:1649-1658(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RX STRAIN=Sprague-Dawley;
RC MEDLINE=21024193; PubMed=11149472;
RA Thunyakitpisal P., Alvarez M., Tokunaga K., Onyia J.E., Hock J.,

```

```

RA Chashi N., Feister H., Rhodes S.J., Bidwell J.P.;
RT "Cloning and functional analysis of a family of nuclear matrix
RT transcription factors (NP/NMP4) that regulate type I collagen
RT expression in osteoblasts.";
RL J. Bone Miner. Res. 16:10-23(2001).
CC -1- FUNCTION: Transcription factor that binds the consensus DNA
CC sequence [GC]AAAAA. Seems to bind and regulate the promoters of
CC MMP1, MMP3, MMP7 and COL1A1.
CC -1- SUBUNIT: Interacts with Cas.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q9EQJ4-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9EQJ4-2; Sequence=VSP_006921;
CC Name=3;
CC IsoId=Q9EQJ4-3; Sequence=VSP_006922;
CC -1- TISSUE SPECIFICITY: Expressed in osteocytes, osteoblasts, and
CC chondrocytes in bone.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: Contains 8 C2H2-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB019281; BAA89664.1; -
DR EMBL; AF216804; AAG40582.1; -
DR EMBL; AF216805; AAG40583.1; -
DR EMBL; AF216806; AAG40584.1; -
DR HSSP; P08153; 1ZFD.
DR TRANSFAC; T05136; -
DR TRANSFAC; T05137; -
DR TRANSFAC; T05138; -
DR TRANSFAC; T05141; -
DR TRANSFAC; T05142; -
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR ProDom; PD000003; Znf_C2H2; 4.
DR SMART; SM00355; Znf_C2H2; 8.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 8.
KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;
KW DNA-binding; Repeat; Alternative splicing.
FT ZN_FING 229 251 C2H2-TYPE 1.
FT ZN_FING 257 279 C2H2-TYPE 2.
FT ZN_FING 285 307 C2H2-TYPE 3.
FT ZN_FING 318 340 C2H2-TYPE 4.
FT ZN_FING 346 368 C2H2-TYPE 5.
FT ZN_FING 374 398 C2H2-TYPE 6.
FT ZN_FING 404 426 C2H2-TYPE 7.
FT ZN_FING 434 456 C2H2-TYPE 8.
FT DOMAIN 462 524 GLN-RICH.
FT DOMAIN 467 506 ALA-RICH.
FT VARSPLIC 103 118 Missing (in isoform 2).
FT VARSPLIC 301 361 Missing (in isoform 3).
FT VARSPLIC 178 179 Missing (in isoform 3).
FT CONFLICT 178 179 GG -> RS (IN REF. 1).
FT CONFLICT 576 577 LA -> WP (IN REF. 1).
SQ SEQUENCE 579 AA; 63139 MW; PBC242E0D1050C45 CRC64;
Query Match 78.0%; Score 39; DB 1; Length 579;
Best Local Similarity 85.7%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      2 PRPPRGR 8
Db      189 PRPPRGR 195

RESULT 6
MM24_MOUSE
ID MM24_MOUSE STANDARD; PRT; 618 AA.
AC Q9R0S2; Q920J9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Matrix metalloproteinase-24 precursor (EC 3.4.24.-) (MMP-24)
DE (Membrane-type matrix metalloproteinase 5) (MT-MMP 5) (Membrane-type-5
DE matrix metalloproteinase) (MT5-MMP) (MMP-21).
GN MMP24 OR MT5MMP OR MMP21.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Seiki M.;
RT "Identification of a new membrane-type matrix metalloproteinase, MT5-
RT MMP, that is expressed predominantly in cerebellum."
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A., AND MUTAGENESIS OF GLU-256.
RC STRAIN-BALB/C; TISSUE=Brain;
RX MEDLINE=99185121; PubMed=10085137;
RA Pei D.Q.;
RT "Identification and characterization of the fifth membrane-type matrix
RT metalloproteinase MT5-MMP."
RL J. Biol. Chem. 274:8925-8932(1999).
RN [3]
RP FUNCTION.
RX MEDLINE=20086420; PubMed=10622708;
RA Wang X., Yi J., Lei J., Pei D.Q.;
RT "Expression, purification and characterization of recombinant mouse
RT MT5-MMP protein products."
RL FEBS Lett. 462:261-266(1999).
CC -!- FUNCTION: Activates progelatinase A. May also be a proteoglycanase
CC involved in degradation of proteoglycans, such as dermatan sulfate
CC and chondroitin sulfate proteoglycans. Cleaves partially
CC fibronectin, but not collagen type I, nor laminin.
CC -!- CORFACTOR: Binds 1 zinc ion per subunit, calcium (By similarity).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SHED FROM CELL
CC SURFACE AS SOLUBLE PROTEINASE, BY A PROTEOLYTIC CLEAVAGE.
CC -!- TISSUE SPECIFICITY: Expressed in brain. Expressed at low level in
CC testis.
CC -!- DEVELOPMENTAL STAGE: Expressed at day 11 until day 15, before
CC dropping around day 17 before birth.
CC -!- PTM: the precursor is cleaved by a furin endopeptidase (By
CC similarity).
CC -!- SIMILARITY: Belongs to peptidase family M10A.
CC -!- SIMILARITY: Contains 1 hemopexin-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB012265; BAA82966.1; -
CC EMBL; AJ010262; CAA09055.1; -
CC HSPS; P03956; 1CGL.
CC MEROPS; M10.023; -
CC MGD; MGI:1341867; Mmp24.
CC InterPro; IPR000585; Hemopexin.
CC InterPro; IPR001818; Pept_M10A_M12B.
CC InterPro; IPR006025; Pept_M_Zn_BS.

```

InterPro; IPR006026; Peptidase\_M.  
Pfam; PF00045; hemopexin; 4.  
Pfam; PF00413; Peptidase\_M10; 1.  
Pfam; PF03933; Peptidase\_M10\_N; 1.  
PRINTS; PR00138; MATRIXIN.  
SMART; SM00120; HX; 4.  
SMART; SM00235; ZnMG; 1.  
PROSITE; PS00546; CYSTEINE\_SWITCH; FALSE\_NEG.  
PROSITE; PS00024; HEMOPEXIN; 1.  
PROSITE; PS00142; ZINC\_PROTEASE; 1.  
KW Hydrolase; Metalloprotease; Zinc; Calcium; Signal; Zymogen;  
KW Transmembrane; Extracellular matrix.  
FT SIGNAL 1 41  
FT PROPEP 42 128  
FT CHAIN 129 618  
FT DOMAIN 42 575  
FT TRANSMEM 576 596  
FT DOMAIN 597 618  
FT DOMAIN 350 545  
FT SITE 112 112  
FT METAL 255 255  
FT ACT\_SITE 256 256  
FT METAL 259 259  
FT METAL 265 265  
FT DISULFD 353 542  
FT DOMAIN 122 125  
FT CONFLICT 7 28  
FT CONFLICT 44 50  
FT CONFLICT 306 308  
FT CONFLICT 326 326  
FT CONFLICT 337 341  
FT CONFLICT 449 449  
FT CONFLICT 502 502  
FT CONFLICT 589 589  
SQ SEQUENCE 618 AA; 70490 MW; 62C0086E1E54B106 CRC64;

Query Match 78.0%; Score 39; DB 1; Length 618;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRPFR 6  
Db 331 HRPFR 336

RESULT 7  
ACVS\_NOCILA  
ID ACVS\_NOCILA STANDARD; PRT; 3649 AA.  
AC P27743;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-AUG-1992 (Rel. 23, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE N-(5-amino-5-carboxypentanoyl)-L-cysteinyl-D-valine synthase  
DE (EC 6.3.2.26) (Delta-(L-alpha-aminoadipyl)-L-cysteinyl-D-valine  
DE synthetase) (ACV synthetase) (ACVS).  
GN PCBA.  
OS Nocardia lactamdurans.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Pseudonocardiaceae; Pseudonocardiaceae; Amycolatopsis.  
OX NCBI\_TaxID=1913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VAR JC 411;  
RX MEDLINE=92065808; PubMed=1956290;  
RA Cogue J.J.R., Martin J.F., Calzada J.G., Liras P.;  
RT "The cephamycin biosynthetic genes pcbaB, encoding a large  
RT multidomain peptide synthetase, and pcBC of Nocardia lactamdurans are  
RT clustered together in an organization different from the same genes  
RT in Actinonidium chrysogenum and Penicillium chrysogenum.";  
RL Mol. Microbiol. 5:1125-1133(1991).  
CC -!- FUNCTION: Each of the constituent amino acids of the tripeptide  
CC acv are activated as aminoacyl-adenylates with peptide bonds



```
RESULT 9
DPM1_CRIGR STANDARD; PRT; 266 AA.
AC Q9WU83;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dolichol-phosphate mannosyltransferase (EC 2.4.1.83) (Dolichol-
DE phosphate mannosase synthase) (Dolichyl-phosphate beta-D-
DE mannosyltransferase) (Mannose-P-dolichol synthase) (MPD synthase) (DPM
DE synthase).
GN DPM1.
OS Cricetulus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RA Pu L., Socca J.R., Walker B.K., Wu J.S., Krag S.S.;
RT "Mutation in B4-2-1 CHO cells defective in MPD synthase activity.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Transfers mannose from GDP-mannose to dolichol
CC monophosphate to form dolichol phosphate mannose (Dol-P-Man) which
CC is the mannosyl donor in pathways leading to N-glycosylation,
CC glycosyl phosphatidylinositol membrane anchoring, and O-
CC mannosylation of proteins (By similarity).
CC -!- CATALYTIC ACTIVITY: GDP-mannose + dolichyl phosphate = GDP +
CC dolichyl D-mannosyl phosphate.
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum (By similarity).
CC -!- SIMILARITY: Belongs to the Glycosyltransferase family 2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF121895; AAD30975.1; -.
CC InterPro: IPR001173; Glyco trans 2.
CC Pfam: PF05335; Glycos_transf_2; 1.
CC Transferrase; Glycosyltransferase; Endoplasmic reticulum.
CC SEQUENCE 266 AA; 29654 MW; 4AFB37EA3AC3329D CRC64;
CC
CC Query Match 76.0%; Score 38; DB 1; Length 266;
CC Best Local Similarity 85.7%; Pred. No. 25;
CC Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 2 PRPPRGR 8
CC |||||
CC Db 20 PRPPQGR 26
CC
CC RESULT 10
ARCC_FAT STANDARD; PRT; 92 AA.
AC P36576;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Arrestin-C (Cone arrestin) (Fragment).
DE ARR3 OR CAR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Sprague-Dawley; TISSUE=Pineal gland;
RC MEDLINE=94140898; PubMed=8308033;
RX
```

```
RA Craft C.M., Whitmore D.H., Wiechmann A.F.;
RT "Cone arrestin identified by targeting expression of a functional
RT family.";
RL J. Biol. Chem. 269:4613-4619(1994).
CC -!- FUNCTION: May play a role in an as yet undefined retina-specific
CC signal transduction. Could binds to photoactivated-phosphorylated
CC red/green opsins.
CC -!- TISSUE SPECIFICITY: Retina and pineal gland.
CC -!- SIMILARITY: Belongs to the arrestin family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U03628; AAA17552.1; -.
CC PIR: I70113; I70113.
CC HSP: P08168; ICF1.
CC InterPro: IPR000698; Arrestin.
CC InterPro: IPR007110; Ig-like.
CC Pfam: PF02752; arrestin_C; 1.
CC ProDom: PD002099; Arrestin; 1.
CC PROSITE; PS00295; ARRESTINS; PARTIAL.
CC Sensory transduction; Vision.
CC -!- NON TER 1
CC SEQUENCE 92 AA; 9878 MW; 88F0C948643C83B9 CRC64;
CC
CC Query Match 74.0%; Score 37; DB 1; Length 92;
CC Best Local Similarity 71.4%; Pred. No. 12;
CC Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC
CC QY 1 HPRPRRG 7
CC |||||
CC Db 61 HPRPDHG 67
CC
CC RESULT 11
YMU5_STRCM STANDARD; PRT; 237 AA.
AC Q05071;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 25.0 kDa protein in muta 5' region (ORF-D).
DE Streptomyces cinnamonensis.
DE Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
DE Streptomycinae; Streptomycetaceae; Streptomyces.
DE NCBI_TaxID=1900;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=A3823.5;
RX MEDLINE=93273720; PubMed=8099072;
RA Birch A., Leiser A., Robinson J.A.;
RT "Cloning, sequencing, and expression of the gene encoding
RT methylmalonyl-coenzyme A mutase from Streptomyces cinnamonensis.";
RL J. Bacteriol. 175:3511-3519(1993).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L10064; AAA03039.1; -.
CC PIR: D40595; D40595.
CC Hypothetical protein.
CC SEQUENCE 237 AA; 25025 MW; D9C55FED795A8455 CRC64;
CC
```

```
Query Match          74.0%; Score 37; DB 1; Length 237;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 PRPPRG 7
DB      113 PRPPRG 118

RESULT 12
NEF_HV2RO
ID -NEF_HV2RO      STANDARD;      PRT;      256 AA.
AC P04500;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Negative factor (F-protein) (27 kDa protein) (3'ORF).
GN NEF.
OS Human immunodeficiency virus type 2 (isolate ROD) (HIV-2).
OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11720;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87173056; PubMed=3031510;
RA Guyader M., Emerman M., Sonigo P., Clavel F., Montagnier L.,
RA Alizon M.;
RT "Genome organization and transactivation of the human
RT immunodeficiency virus type 2.";
RL Nature 326:662-669(1987).
RN [2]
RP POST-TRANSLATIONAL MODIFICATIONS, AND FUNCTION.
RX MEDLINE=88039140; PubMed=3118220;
RA Guy B., Kiely M.-P., Riviere Y., le Peuch C., Dott K., Girard M.,
RA Montagnier L., Lecocq J.-P.;
RT "HIV F/3' orf encodes a phosphorylated GTP-binding protein resembling
RT an oncogene product.";
RL Nature 330:266-269(1987).
CC -!- FUNCTION: NEF has GTPase, GTP-binding and autophosphorylating
CC activities. It seems to down-regulate the CD4(T4) antigen.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M15390; AAB00771.1; -
CC F01; D28262; ASLJH2.
CC HIV; M15390; NEFS2ROD.
CC InterPro; IPR001558; HIV_Nef.
CC Pfam; PF00469; F-protein; 1.
CC ProDom; PD000031; HIV_Nef; 1.
CC AIDS; Myristate; GTP-binding; phosphorylation; Lipoprotein.
FT LIPID 2 2 N-myristoyl glycine (in host).
FT MOD_RES 10 10 PHOSPHORYLATION (BY PKC).
SQ SEQUENCE 256 AA; 29566 MW; 55BA1B7871CCAB5 CRC64;

Query Match          74.0%; Score 37; DB 1; Length 256;
Best Local Similarity 85.7%; Pred. No. 34;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 HRPFRPG 7
DB      9 HSRPPRG 15

RESULT 13
ERMA_ARTS3
ID -ERMA_ARTS3     STANDARD;      PRT;      340 AA.
AC P09891;

Query Match          74.0%; Score 37; DB 1; Length 340;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 PRPPRG 7
DB      332 PRPPRG 337

RESULT 14
GAG_HTL1A
ID -GAG_HTL1A      STANDARD;      PRT;      429 AA.
AC P03345;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE GAG polyprotein [Contains: Major core proteins P19 and P24; Nucleic
DE acid-binding protein p15].
GN GAG.
OS Human T-cell leukemia virus type I (strain ATK) (HTLV-I).
OC Viruses; Retroviral viruses; Retroviridae; Deltaretrovirus.
OX NCBI_TaxID=11926;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83221647; PubMed=6304725;
RA Seiki M., Hattori S., Hirayama Y., Yoshida M.;
RT "Human adult T-cell leukemia virus: complete nucleotide sequence of
RT the provirus genome integrated in leukemia cell DNA.";
RT Proc. Natl. Acad. Sci. U.S.A. 80:3618-3622(1983).
```

```
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE RNA adenine N-6-methyltransferase (EC 2.1.1.48) (Macrolide-
DE lincosamide-streptogramin B resistance protein) (Erythromycin
DE resistance protein).
DE ERMA.
GN Arthrobacter sp. (strain B3381).
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Propionibacteriaceae; Nocardioidaceae; Aeromicrobium.
OX NCBI_TaxID=31956;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86006275; PubMed=4043733;
RA Roberts A.N., Hudson G.S., Brenner S.;
RT "An erythromycin-resistance gene from an erythromycin-producing
RT strain of Arthrobacter sp.";
RL Gene 35:259-270(1985).
CC -!- FUNCTION: THIS PROTEIN PRODUCES A DIMETHYLATION OF THE ADENINE
CC RESIDUE AT POSITION 2058 IN 23S RNA, RESULTING IN REDUCED
CC AFFINITY BETWEEN RIBOSOMES & MACROLIDE-LINCOSAMIDE-STREPTOGRAMIN B
CC ANTIBIOTICS.
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + rRNA = S-adenosyl-L-
CC homocysteine + rRNA containing N(6)-methyladenine.
CC -!- SIMILARITY: Belongs to the rRNA adenine N-6-methyltransferase
CC family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M11276; AAA22075.1; -
CC InterPro; IPR001737; RNA_A_dimeth.
CC InterPro; IPR000051; SAM_Bind.
CC Pfam; PF00398; RnaAD; 1.
CC SMART; SM00650; RABC; 1.
CC PROSITE; PS01131; RNA_A_DIMETH; 1.
KW Antibiotic resistance; Transferase; Methyltransferase.
SQ SEQUENCE 340 AA; 37453 MW; E99A714C391952B5 CRC64;

Query Match          74.0%; Score 37; DB 1; Length 340;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 PRPPRG 7
DB      332 PRPPRG 337
```

[2]
RN SEQUENCE OF 131-155.
RX MEDLINE=82174582; PubMed=6280175;
RA Oroszlan S., Saragadharan M.G., Copeland T.D., Kalyanaraman V.S.,
RT "Primary structure analysis of the major internal protein p24 of
RT human type C-T-cell leukemia virus";
RL Proc. Natl. Acad. Sci. U.S.A. 79:1291-1294 (1982).
[3]
RN SEQUENCE OF 345-429.
RX MEDLINE=84029174; PubMed=6313426;
RA Copeland T.D., Oroszlan S., Kalyanaraman V.S., Saragadharan M.G.,
RT "Complete amino acid sequence of human T-cell leukemia virus
RT structural protein p15";
RL FEBS Lett. 162:390-395 (1983).
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -!- MISCELLANEOUS: This protein is synthesized as a Gag-Pol
CC polypeptide.
CC -!- SIMILARITY: Contains 2 CCHC-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J02029; AAA96672.1; -.
CC PIR; B93954; FOLJGH.
CC PDB; 1ORJ; 26-SEP-01.
CC InterPro; IPR003139; Gag\_p19.
CC InterPro; IPR000721; Gag\_p24.
CC InterPro; IPR008916; Retrov\_capsid\_C.
CC InterPro; IPR001878; Znf\_CCHC.
CC InterPro; IPR008919; Retrov\_capsid\_N.
CC Pfam; PF02228; Gag\_p19; 1.
CC Pfam; PF00607; Gag\_p24; 1.
CC Pfam; PF00098; Zf\_CCHC; 2.
CC PRINTS; PR00939; C2HCZNFINGER.
CC SMART; SM00343; Znf\_CCHC; 2.
CC PROSITE; PS50158; Zf\_CCHC; 1.
CC Core protein; Polyprotein; Zinc-finger; Repeat.
CC CHAIN 1 130 MAJOR CORE PROTEIN P19.
CC CHAIN 131 344 MAJOR CORE PROTEIN P24.
CC CHAIN 345 429 NUCLEIC ACID-BINDING PROTEIN P15.
CC ZN\_FING 378 395 CCHC-TYPE 1.
CC ZN\_FING 395 395 CCHC-TYPE 2.
CC A -> T (IN REF. 2). AAA96672).
CC S -> F (IN REF. 1; AAA96672).
CC CONFLICT 416 416
CC CONFLICT 424 424
CC CONFLICT 424 424
CC SEQUENCE 429 AA; 47496 MW; 03FFD4E5A4500284 CRC64;
CC
CC Query Match 74.0%; Score 37; DB 1; Length 429;
CC Best Local Similarity 100.0%; Pred. No. 59;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 2 PRPPRG 7
CC Db 13 PRPPRG 18
CC
CC RESULT 15
CC GAG\_HTL1C STANDARD; PRT; 429 AA.
CC AC P14076;
CC DT 01-JAN-1990 (Rel. 13, Created)
CC DT 01-JAN-1990 (Rel. 13, Last sequence update)
CC DT 28-FEB-2003 (Rel. 41, Last annotation update)
CC DE GAG Polyprotein [Contains: Major core proteins p19 and p24; Nucleic
CC acid-binding protein p15].
CC GN GAG.
CC OS Human T-cell leukemia virus type I (Caribbean isolate) (HTLV-I).

Viruses; Retrovirdae; Retroviridae; Deltaretrovirus.
NX NCB1\_TaxID=11927;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=88274338; PubMed=2899128;
RA Malik K.T.A., Even J., Karpas A.;
RT "Molecular cloning and complete nucleotide sequence of an adult T
RT cell leukaemia virus/human T cell leukaemia virus type I
RT (ATLV/HTLV-I) isolate of Caribbean origin: relationship to other
RT members of the ATLV/HTLV-I subgroup.";
RL J. Gen. Virol. 69:1695-1710 (1988).
CC -!- PTM: Specific enzymatic cleavages in vivo yield mature proteins.
CC -!- SIMILARITY: Contains 2 CCHC-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D13784; BAA02929.1; -.
CC PIR; A28136; FOLJCN.
CC InterPro; IPR003139; Gag\_p19.
CC InterPro; IPR000721; Gag\_p24.
CC InterPro; IPR008916; Retrov\_capsid\_C.
CC InterPro; IPR008919; Retrov\_capsid\_N.
CC InterPro; IPR001878; Znf\_CCHC.
CC Pfam; PF02228; Gag\_p19; 1.
CC Pfam; PF00607; Gag\_p24; 1.
CC Pfam; PF00098; Zf\_CCHC; 2.
CC PRINTS; PR00939; C2HCZNFINGER.
CC SMART; SM00343; Znf\_CCHC; 2.
CC PROSITE; PS50158; Zf\_CCHC; 1.
CC Core protein; Polyprotein; Zinc-finger; Repeat.
CC CHAIN 1 130 MAJOR CORE PROTEIN P19.
CC CHAIN 131 344 MAJOR CORE PROTEIN P24.
CC CHAIN 345 429 NUCLEIC ACID-BINDING PROTEIN P15.
CC ZN\_FING 378 395 CCHC-TYPE 1.
CC ZN\_FING 395 395 CCHC-TYPE 2.
CC SEQUENCE 429 AA; 47514 MW; CFBF686497B738EC CRC64;
CC
CC Query Match 74.0%; Score 37; DB 1; Length 429;
CC Best Local Similarity 100.0%; Pred. No. 59;
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC QY 2 PRPPRG 7
CC Db 13 PRPPRG 18
CC
CC Search completed: April 6, 2004, 16:08:00
CC Job time : 5.15888 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:51:34 ; Search time 28.9346 Seconds  
(without alignments)  
87.236 Million cell updates/sec

Title: US-10-009-709-9  
Perfect score: 50  
Sequence: 1 HPRPRGR 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: SPTREMBL\_25.\*
- 2: sp\_archaea.\*
- 3: sp\_bacteria.\*
- 4: sp\_fungi.\*
- 5: sp\_human.\*
- 6: sp\_invertebrate.\*
- 7: sp\_mammal.\*
- 8: sp\_mhc.\*
- 9: sp\_organelle.\*
- 10: sp\_phage.\*
- 11: sp\_plant.\*
- 12: sp\_rodent.\*
- 13: sp\_virus.\*
- 14: sp\_vertebrate.\*
- 15: sp\_unclassified.\*
- 16: sp\_rvirus.\*
- 17: sp\_bacteriap.\*
- 18: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	90.0	407	10	Q8H3U2
2	44	88.0	561	16	O05316
3	44	88.0	561	16	Q7U0D5
4	43	86.0	545	2	O05089
5	42	84.0	332	2	Q54729
6	42	84.0	494	5	Q8T777
7	42	84.0	535	10	Q84SD0
8	41	82.0	633	10	Q7X188
9	41	82.0	2796	2	Q48926
10	40	80.0	129	17	O57931
11	39	78.0	200	12	Q66619
12	39	78.0	216	10	Q9C591
13	39	78.0	238	10	Q9PC5
14	39	78.0	246	16	Q8X1J6
15	39	78.0	256	10	Q84J66
16	39	78.0	268	10	Q39487

17	39	78.0	346	11	Q60535
18	39	78.0	397	11	Q9EQJ1
19	39	78.0	460	4	Q7Z722
20	39	78.0	470	16	P95042
21	39	78.0	470	16	Q7U1F9
22	39	78.0	517	11	Q8C3E0
23	39	78.0	526	2	Q8VPR7
24	39	78.0	559	10	Q943J0
25	39	78.0	602	5	Q9GN10
26	38	76.0	157	16	Q9RXJ9
27	38	76.0	264	10	Q7V1M9
28	38	76.0	299	10	Q85261
29	38	76.0	450	16	Q8U651
30	38	76.0	632	2	Q9KWF1
31	38	76.0	884	10	Q94J11
32	38	76.0	2406	4	Q9BZS0
33	38	76.0	2414	4	Q9HCL7
34	37	74.0	72	15	Q86918
35	37	74.0	107	17	Q9VD88
36	37	74.0	117	16	Q9KZ63
37	37	74.0	118	6	Q9XSS0
38	37	74.0	123	17	Q9YAM7
39	37	74.0	129	15	Q9IZL0
40	37	74.0	136	4	Q9NWG6
41	37	74.0	137	5	Q45200
42	37	74.0	151	16	Q9PAK1
43	37	74.0	163	16	Q8VU20
44	37	74.0	189	5	Q97409
45	37	74.0	199	5	Q816P8

ALIGNMENTS

RESULT 1

Q8H3U2 PRELIMINARY; PRT; 407 AA.

AC Q8H3U2;  
DT 01-MAR-2003 (TREMELrel. 23, Created)  
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)  
DT 01-MAR-2003 (TREMELrel. 23, Last annotation update)  
DE P0418E08.7 protein.  
GN P0418E08.7.  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC SFRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC  
clone:P0418E08.";  
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF004382; BAC16073.1; -  
SQ SEQUENCE 407 AA; 43973 MW; D39E5777F8252C8 CRC64;

Query Match 90.0%; Score 45; DB 10; Length 407;  
Best Local Similarity 87.5%; Pred No. 11;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HPRPRGR 8  
Db 118 HPEPRGR 125

RESULT 2

O05316 PRELIMINARY; PRT; 561 AA.

ID O05316;  
AC O05316;  
DT 01-JUL-1997 (TREMELrel. 04, Created)  
DT 01-JUL-1997 (TREMELrel. 04, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein (Diester hydrolase, putative).  
 GN RV1215C OR MTQ1364.27C OR MT1253.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 Gordon S.V., Eigemeier K., Gas S., Barry C.E. III, Tekala F.,  
 Badcock K., Basham D., Brown D., Chillingworth I., Connor R.,  
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 Oliver S., Osborne K., Quail M.A., Rajandream M.A., Rogers J.,  
 Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 Sulston J.E., Taylor K., Whitehead S., Barrall B.G.,  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 Peterson J.F., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 Kolonay J.F., Nelson W.C., Oram L.A., Ermolaeva M.D., Salzberg S.L.,  
 Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 Bishai W.;  
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DDAJ databases.  
 DR EMBL; Z93777; CAB07817.1; --  
 DR EMBL; AE007002; AA45510.1; --  
 DR PIR; E70610; E70610.  
 DR TIGR; MT1253; --  
 DR TuberculList; RV1215C; --  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0008451; F:1xaa-pro aminopeptidase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR008979; Gal Bind like.  
 DR InterPro; IPR001005; Myb DNA Binding.  
 DR InterPro; IPR000383; Peptidase S15.  
 DR Pfam; PF01219; Peptidase\_S15; 1.  
 DR TIGRFAMs; TIGR00976; /NonD; 1.  
 DR PROSITE; PS00037; MYB\_1; 1.  
 KW Hypothetical protein; Hydrolase; Complete proteome.  
 SQ SEQUENCE 561 AA; 62610 MW; 6D2C3253F2D3598D CRC64;  
 Query Match 88.0%; Score 44; DB 16; Length 561;  
 Best Local Similarity 87.5%; Pred. No. 22;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HPRPPRGR 8  
 DB 534 YPRPPRGR 541  
 RESULT 3  
 Q7U0D5 PRELIMINARY; PRT; 561 AA.  
 AC Q7U0D5;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Conserved hypothetical protein.  
 GN MB1247C.  
 OS Mycobacterium bovis.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AF2122/97;  
 RX MEDLINE=22703107; PubMed=12788972;  
 RA Garnier T., Eigemeier K., Camus J.-C., Medina N., Mansoor H.,  
 Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,  
 Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
 RA "The complete genome sequence of Mycobacterium bovis."  
 RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).  
 RL EMBL; EX248338; CAD94108.1; --  
 KW Complete proteome.  
 SQ SEQUENCE 561 AA; 62626 MW; C35176B8172866AD CRC64;  
 Query Match 88.0%; Score 44; DB 16; Length 561;  
 Best Local Similarity 87.5%; Pred. No. 22;  
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 HPRPPRGR 8  
 DB 534 YPRPPRGR 541  
 RESULT 4  
 O55089 PRELIMINARY; PRT; 545 AA.  
 AC O55089;  
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
 DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Nocardioides simplex (Arthrobacter simplex).  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Propionibacterineae; Nocardioidaceae; Pimelobacter.  
 OX NCBI\_TaxID=2045;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IFO12069;  
 RX MEDLINE=95319311; PubMed=7596291;  
 RA Molnar I., Choi K., Yamashita M., Murooka Y.;  
 RT "Molecular cloning, expression in Streptomyces lividans, and analysis  
 RT of a gene cluster from Arthrobacter simplex encoding 3-  
 RT ketosteroid- $\Delta^1$ -dehydrogenase, 3-ketosteroid- $\Delta^1$ -isomerase  
 RT and a hypothetical regulatory protein."  
 RL Mol. Microbiol. 15:895-905 (1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=IFO12069;  
 RA Dziadek J., Yamashita M., Murooka Y.;  
 RT "Cloning, sequencing and characterization of the downstream region of  
 RT ksdI operon of Arthrobacter simplex."  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z93338; CA807541.1; --  
 KW Hypothetical protein.  
 SQ SEQUENCE 545 AA; 57318 MW; 55F441EBD5E3A19E CRC64;  
 Query Match 86.0%; Score 43; DB 2; Length 545;  
 Best Local Similarity 87.5%; Pred. No. 32;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 HPRPPRGR 8  
 DB 456 HPPPPRGR 463  
 RESULT 5  
 Q54729 PRELIMINARY; PRT; 332 AA.  
 ID Q54729  
 AC Q54729;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)



```

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE ORF11692 (Fragment).
OS Synecococcus sp. (strain PCC 7942) (Anacystis nidulans R2).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OC NCBI_TaxID=1140;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCC 7942;
RA Tsinoirenas N.F.; Golden S.S.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U44761; AAA86647.1; -.
DR InterPro; IPR001064; Crystallin.
DR PROSITE; PS00225; CRYSTALLIN_BTAGMWA; 1.
FT NON_TER 1
SQ SEQUENCE 332 AA; 38574 MW; CA17B5F4B0F2ED09 CRC64;

Query Match      84.0%; Score 42; DB 2; Length 332;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PRPPRGR 8
DB 282 PRPPRGR 288

RESULT 6
O8T777 PRELIMINARY; PRT; 494 AA.
AC Q8T777;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Branchiostoma floridae (Florida lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OC NCBI_TaxID=7739;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=11967531.
RA Abi-Rached L.; Gilles A.; Shih A.; Pontarotti P.; Inoko H.;
RL Nat. Genet. 0:0-0(2002).
RT "Evidence of en bloc duplication in vertebrate genomes."
DR EMBL; AF391288; AAM18867.1; -.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR008922; Di-copper_centre.
DR InterPro; IPR002227; Tyrosinase.
DR Pfam; PF00264; Tyrosinase; 1.
DR PRINTS; PR00092; TYROSINASE.
DR PROSITE; PS00497; TYROSINASE_1; 1.
DR PROSITE; PS00498; TYROSINASE_2; 1.
KW Hypothetical protein.
SQ SEQUENCE 494 AA; 56646 MW; 6823A97C0CF1A05E CRC64;

Query Match      84.0%; Score 42; DB 5; Length 494;
Best Local Similarity 87.5%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HPPRPPGR 8
DB 30 HPPRPPGR 37

RESULT 7
Q84SD0 PRELIMINARY; PRT; 535 AA.
ID Q84SD0;
AC Q84SD0;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE P0577B11.16 protein.

```

```

GN P0577B11.16.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OC NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T.; Matsumoto T.; Katayose Y.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, PAC
clone:P0577B11."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005504; BAC57819.1; -.
DR InterPro; IPR002885; PPR.
DR Pfam; PF01535; PPR; 8.
DR TIGRPFAMs; TIGR00756; PPR; 9.
SQ SEQUENCE 535 AA; 58557 MW; C2712C4178582982 CRC64;

Query Match      84.0%; Score 42; DB 10; Length 535;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PRPPRGR 8
DB 114 PRPPRGR 120

RESULT 8
Q7X188 PRELIMINARY; PRT; 633 AA.
ID Q7X188;
AC Q7X188;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Zinc finger protein-like protein.
GN P0453G03.32.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzeae; Oryza.
OC NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T.; Matsumoto T.; Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, PAC
clone:P0453G03."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004276; BAC79830.1; -.
SQ SEQUENCE 633 AA; 67636 MW; EDB82E36082A61F0 CRC64;

Query Match      82.0%; Score 41; DB 10; Length 633;
Best Local Similarity 75.0%; Pred. No. 78;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HPPRPPGR 8
DB 497 HPPRPPGR 504

RESULT 9
Q48926 PRELIMINARY; PRT; 2796 AA.
ID Q48926;
AC Q48926;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Fatty acid synthase.
GN FAS.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

```

```

CX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BCG-Pasteur;
RX MEDLINE=96200863; PubMed=8621098;
RA Fernandes N.D., Kolattukudy P.E.;
RT "Cloning, sequencing and characterization of a fatty acid synthase-
  encoding gene from Mycobacterium tuberculosis var. bovis BCG.";
PL Gene 170.95-99(1996);
DR EMBL; U36763; AAB03809.1; -.
DR FIRM; JC4743; JC4743.
DR GO; GO:0005835; C:fatty-acid synthase complex; IEA.
DR GO; GO:0004312; F:fatty-acid synthase activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR GO; GO:0006333; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR001227; Ac trans.
DR InterPro; IPR003965; Fatty acid synth.
DR InterPro; IPR00794; Ketoacyl synth.
DR InterPro; IPR002539; MaoC dehydratase.
DR Pfam; PF00698; Acyl_transf; 1.
DR Pfam; PF00109; ketoacyl-synt; 1.
DR Pfam; PF02801; ketoacyl-synt C; 1.
DR Pfam; PF01575; MaoC dehydratase; 1.
DR PRINTS; PR01483; FASYNTHASE.
DR PROSITE; PS00606; B_KETOACYL SYNTHASE; 1.
SQ SEQUENCE 2796 AA; 303675 MW; 47887169126D2504 CRC64;

Query Match 82.0%; Score 41; DB 2; Length 2796;
Best Local Similarity 87.5%; Pred. No. 3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HRPFRGR 8
Db 422 HRPFRGR 429
|||||

RESULT 10
ID O57931 PRELIMINARY; PRT; 129 AA.
AC O57931;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein PH0192.
GN PH0192.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococci; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Koeugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
  thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA RES. 5:55-76(1998);
DR EMBL; AP000001; BAA29261.1; -.
DR FIRM; F71241; F71241.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 129 AA; 14331 MW; B5E1E49F0F8F2F1 CRC64;

Query Match 80.0%; Score 40; DB 17; Length 129;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 1 HRPFRGR 7
Db 83 HRPFRGR 89
|||||

RESULT 11
ID Q66619 PRELIMINARY; PRT; 200 AA.
AC Q66619;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE ORF 13.
OS Equine herpesvirus type 2 (strain 86/87) (EHV-2).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae.
OX NCBI_TaxID=82831;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95302501; PubMed=7783207;
RA Telford E.A., Watson M.S., Aird H.C., Perry J., Davison A.J.;
RT "The DNA sequence of equine herpesvirus 2.";
RL J. Mol. Biol. 249:520-528(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Telford E.A.R.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20824; AAC13802.1; -.
DR FIRM; S55609; S55609.
SQ SEQUENCE 200 AA; 22356 MW; CF3306CF3C0974B1 CRC64;

Query Match 78.0%; Score 39; DB 12; Length 200;
Best Local Similarity 85.7%; Pred. No. 58;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 PPRPRGR 8
Db 13 PPRPRGR 19
|||||

RESULT 12
ID Q9C591 PRELIMINARY; PRT; 216 AA.
AC Q9C591;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein (AP2 domain transcription factor, putative).
GN AT5G21960.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Koeugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
  thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA RES. 5:55-76(1998);
DR EMBL; AP000001; BAA29261.1; -.
DR FIRM; F71241; F71241.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 129 AA; 14331 MW; B5E1E49F0F8F2F1 CRC64;

Query Match 80.0%; Score 40; DB 17; Length 129;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

DR HSSP; 080337; 2GCC.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0005840; C:ribosome; IEA.
DR GO; GO:0003735; F:structural constituent of ribosome; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0006412; P:protein biosynthesis; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR006032; Ribosomal_S12_23.
DR InterPro; IPR001471; TF_ERF.
DR Pfam; PF00847; AP2-domain; 1.
DR PRINTS; PR00367; ETRHSPELENT.
DR ProDom; PD001423; TF_ERF; 1.
DR SMART; SM00380; AP2; 1.
DR PROSITE; PS00055; RIBOSOMAL_S12; 1.
KW Hypothetical protein.
SQ SEQUENCE 216 AA; 24400 MW; E8C317B74759056E CRC64;

Query Match 78.0%; Score 39; DB 10; Length 216;
Best Local Similarity 85.7%; Pred. No. 62;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HRPFRPG 7
DB 100 HPSPPRG 106

RESULT 13
Q9FFC5 PRELIMINARY; PRT; 238 AA.
ID Q9FFC5
AC Q9FFC5;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE P0679C08.18 protein.
GN P0679C08.18.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 6, PAC
RT clone: P0679C08.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF002542; BAB19378.1; -.
DR Gramene; Q9FFC5; -.
SQ SEQUENCE 238 AA; 26481 MW; BF56639D25433CD CRC64;

Query Match 78.0%; Score 39; DB 10; Length 238;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRPFRP 6
DB 109 HRPFRP 114

RESULT 14
Q8YLJ6 PRELIMINARY; PRT; 246 AA.
ID Q8YLJ6
AC Q8YLJ6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Probable transmembrane protein.
GN RSC0694 OR RS01599.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=305;

```

```

RN SEQUENCE FROM N.A.
RP STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Ariat M., Billault A., Brottier P., Camus J.C., Catolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Sigrier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum.";
RL Nature 415:497-502(2002).
DR EMBL; AL646060; CAD14224.1; -.
DR InterPro; IPR005184; DUF306.
DR Pfam; PF03724; META; 1.
KW Complete proteome.
SQ SEQUENCE 246 AA; 24863 MW; 4127DF6A38274F74 CRC64;

Query Match 78.0%; Score 39; DB 16; Length 246;
Best Local Similarity 75.0%; Pred. No. 70;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HRPFRGR 8
DB 237 HPLPRGQ 244

RESULT 15
Q84J66 PRELIMINARY; PRT; 256 AA.
ID Q84J66
AC Q84J66;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN OSJNBA0015N08.4 OR OSJNBA0096M04.10.
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]_
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.;
RA Fadrosch D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,
RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feldblyum T.V.,
RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
RA White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 3 BAC OSJNBA0015N08 genomic sequence.";
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,
RA Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.;
RA Fadrosch D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,
RA Vanaken S.S., Riedmuller S.B., Utterback T.T., Feldblyum T.V.,
RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,
RA White O., Salzberg S.L., Fraser C.M.;
RT "Oryza sativa chromosome 3 BAC OSJNBA0096M04 genomic sequence.";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC096688; AAO20080.1; -.
DR EMBL; AC092559; AAO37953.1; -.
DR InterPro; IPR001660; SAM.
DR Pfam; PF00536; SAM; 1.
DR SMART; SM00454; SAM; 1.
DR PROSITE; PS50105; SAM_DOMAIN; 1.
KW Hypothetical protein.
SQ SEQUENCE 256 AA; 26728 MW; EC64E19D850D6203 CRC64;

Query Match 78.0%; Score 39; DB 10; Length 256;

```

Best Local Similarity 100.0%; Pred. No. 72;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HPRPPR 6  
| | | | |  
Db 86 HPRPPR 91

Search completed: April 6, 2004, 16:14:42  
Job time : 30.0115 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:29:39 ; Search time 44.785 Seconds  
(without alignments)  
50.472 Million cell updates/sec

Title: US-10-009-709-9

Perfect score: 50

Sequence: 1 HPRPRGR 8

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_29Jan04.\*

1: Geneseq1980s.\*

2: Geneseq1990s.\*

3: Geneseq2000s.\*

4: Geneseq2001s.\*

5: Geneseq2002s.\*

6: Geneseq2003as.\*

7: Geneseq2003bs.\*

8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	8	AAB48779	Human sal
2	50	100.0	9	AAB48778	Human sal
3	50	100.0	10	AAB48777	Human sal
4	50	100.0	17	AAB48783	Human sal
5	50	100.0	132	ABB38848	Peptide #
6	50	100.0	132	AAm32323	Peptide #
7	50	100.0	132	AAm72058	Human bon
8	50	100.0	132	AAm59494	Human bra
9	50	100.0	132	ABG53742	Human liv
10	50	100.0	132	ABG41873	Human pep
11	50	100.0	149	ABR57423	Human NOV
12	50	100.0	154	ABR56769	Human sec
13	50	100.0	166	ADA83798	Human PRH
14	50	100.0	166	ADC98216	Human sal
15	44	88.0	51	AGS58448	Zea may
16	42	84.0	7	AAB48780	Human sal
17	42	84.0	92	ABM65798	Propionib
18	42	84.0	104	ABG18278	Novel hum
19	42	84.0	141	AAU5747	Propionib
20	42	84.0	141	ABM62266	Propionib
21	42	84.0	258	ABG26281	Novel hum
22	42	84.0	306	ABG03818	Novel hum
23	42	84.0	578	ADE07941	Novel pro
24	42	84.0	694	ABG70853	Synechoco
25	42	84.0	19938	ABP76681	Streptomy

#### ALIGNMENTS

##### RESULT 1

AAB48779  
ID AAB48779 standard; peptide; 8 AA.  
AC AAB48779;  
XX  
DT 09-MAR-2001 (first entry)  
XX  
DE Human saliva PRP-1 fragment (residues 101-108), SEQ ID NO:9.  
XX  
KW Human; PRP-1; proline-rich protein; saliva; dental caries;  
KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
KW oral bacterium; caries prevention.  
XX  
OS Homo sapiens.  
XX  
PN W0200069890-A1.  
XX  
PD 23-NOV-2000.  
XX  
PF 11-MAY-2000; 2000MO-SE000930.  
XX  
PR 17-MAY-1999; 99SE-00001773.  
XX  
FA (STRO//) STROEMBERG N.  
FA (JOHA//) JOHANSSON I.  
XX  
PI Stroenberg N, Johansson I;  
XX  
DR WPI; 2001-031923/04.  
XX  
XX  
PT New oligopeptides comprising 2 arginine residues from degradation of  
PT proline-rich proteins, useful for preventing dental caries.  
XX  
PS Claim 4; Page 24; 36pp; English.  
XX  
CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
CC AAB48783) which contain at least two arginine residues and which protect  
CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
CC determinants of a person's susceptibility to dental caries. PRPs are  
CC degraded by Actinomyces and Streptococcus species to small peptide  
CC fragments. These are metabolised by oral bacteria for nutritional  
CC purposes, with certain bacterial species generating ammonia via the  
CC catabolism of arginine. The peptides of the invention, being arginine-  
CC rich, can also be converted to ammonia by these bacteria. The ammonia  
CC thus formed raises the pH at the dental surface, thereby protecting the  
CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-

Abu79133 Mycobacte  
Abp70785 Human ext  
Aay65870 Germline  
Abg14712 Novel hum  
Aau54627 Propionib  
Abm51146 Propionib  
Aau64324 Propionib  
Abm60843 Propionib  
Aac01213 Human pol  
Aau57069 Propionib  
Abm53588 Propionib  
Abb15815 Human ner  
Aag14998 Novel hum  
Aag30036 Arabidops  
Abg14713 Novel hum  
Aac05792 Human pol  
Aau4890 Propionib  
Abm45409 Propionib  
Aam41387 Human pol  
Abu34745 Protein e

26 41 82.0 2796 6 ABU79133  
27 39.5 79.0 140 6 ABP70785  
28 39 78.0 30 3 AAY65870  
29 39 78.0 47 4 ABG14712  
30 39 78.0 50 4 AAU54627  
31 39 78.0 50 6 ABM51146  
32 39 78.0 65 4 AAU64324  
33 39 78.0 65 6 ABM60843  
34 39 78.0 91 4 AAC01213  
35 39 78.0 112 4 AAU57069  
36 39 78.0 112 6 ABM53588  
37 39 78.0 119 4 ABB15815  
38 39 78.0 156 4 AAG14998  
39 39 78.0 216 6 AAG30036  
40 39 78.0 234 4 ABG14713  
41 39 78.0 290 4 AAC05792  
42 39 78.0 308 4 AAU4890  
43 39 78.0 308 6 ABM45409  
44 39 78.0 344 4 AAM41387  
45 39 78.0 470 6 ABU34745

CC derived oligopeptides of the invention

SQ Sequence 8 AA;

Query Match 100.0%; Score 50; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HPRPPRGR 8  
|||||

Db 1 HPRPPRGR 8  
|||||

RESULT 2

ID AAB48778 standard; peptide; 9 AA.

AC AAB48778;

XX AAB48778;

DT 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 100-108), SEQ ID NO:8.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

KW oral bacterium; caries prevention.

XX OS Homo sapiens.

XX WO200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SE000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO//) STROEMBERG N.

XX (JOHA//) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of

PT proline-rich proteins, useful for preventing dental caries.

PS Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
AAB48783) which contain at least two arginine residues and which protect  
CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
CC determinants of a person's susceptibility to dental caries. PRPs are  
CC degraded by Actinomyces and Streptococcus species to small peptide  
CC fragments. These are metabolised by oral bacteria for nutritional  
CC purposes, with certain bacterial species generating ammonia via the  
CC catabolism of arginine. The peptides of the invention, being arginine-  
CC rich, can also be converted to ammonia by these bacteria. The ammonia  
CC thus formed raises the pH at the dental surface, thereby protecting the  
CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
CC derived oligopeptides of the invention

XX SQ Sequence 9 AA;

Query Match 100.0%; Score 50; DB 4; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HPRPPRGR 8  
|||||

Db 2 HPRPPRGR 9  
|||||

RESULT 3

AAB48777 standard; peptide; 10 AA.

XX AAB48777;

AC AAB48777;

XX 09-MAR-2001 (first entry)

DT 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 99-108), SEQ ID NO:7.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

KW oral bacterium; caries prevention.

XX OS Homo sapiens.

XX WO200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SE000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO//) STROEMBERG N.

XX (JOHA//) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of

PT proline-rich proteins, useful for preventing dental caries.

PS Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
AAB48783) which contain at least two arginine residues and which protect  
CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
CC determinants of a person's susceptibility to dental caries. PRPs are  
CC degraded by Actinomyces and Streptococcus species to small peptide  
CC fragments. These are metabolised by oral bacteria for nutritional  
CC purposes, with certain bacterial species generating ammonia via the  
CC catabolism of arginine. The peptides of the invention, being arginine-  
CC rich, can also be converted to ammonia by these bacteria. The ammonia  
CC thus formed raises the pH at the dental surface, thereby protecting the  
CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
CC derived oligopeptides of the invention

XX SQ Sequence 10 AA;

Query Match 100.0%; Score 50; DB 4; Length 10;  
Best Local Similarity 100.0%; Pred. No. 0.38;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HPRPPRGR 8  
|||||

Db 3 HPRPPRGR 10  
|||||

RESULT 4

AAB48783 standard; peptide; 17 AA.

XX AAB48783;

AC AAB48783;

XX 09-MAR-2001 (first entry)

DT Human saliva PRP-1 fragment (residues 99-115), SEQ ID NO:13.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.

XX Homo sapiens.

XX WO2000069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SE000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO/) STROEMBERG N.

XX (JOHA/) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.

XX Claim 2; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention

XX Sequence 17 AA;

Query Match 100.0%; Score 50; DB 4; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 0.61; Indels 0; Gaps 0;  
 Matches 8; Conservative 0; Mismatches 0;

OY 1 HPRPPRGR 8

DB 3 HPRPPRGR 10

RESULT 5

ABB38848  
 ID ABB38848 standard; peptide; 132 AA.

XX ABB38848;

XX 04-FEB-2002 (first entry)

XX Peptide #6354 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000669.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.  
 PR 21-SEP-2000; 2000US-0234687P.  
 PR 27-SEP-2000; 2000US-0236359P.  
 PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human foetal liver.

XX Claim 27; SEQ ID NO 31483; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring  
 CC human gene expression in a sample derived from human foetal liver. The  
 CC single exon nucleic acid probes may be used for predicting, measuring and  
 CC displaying gene expression in samples derived from human foetal liver. The  
 CC present sequence is a peptide encoded by a single exon nucleic acid probe  
 CC of the invention. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 132 AA;

Query Match 100.0%; Score 50; DB 4; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 3.8; Indels 0; Gaps 0;  
 Matches 8; Conservative 0; Mismatches 0;

OY 1 HPRPPRGR 8

DB 83 HPRPPRGR 90

RESULT 6

AAM32323  
 ID AAM32323 standard; protein; 132 AA.

XX AAM32323;

XX 17-OCT-2001 (first entry)

XX Peptide #6360 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000663.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human placenta.

XX Claim 27; SEQ ID NO 32592; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENPs; see AA131315-AA157546). The present sequence is a peptide encoded by one such probe. The probes are useful for producing a microarray for predicting, measuring and displaying gene expression in samples derived from human placenta. The probes are useful for antenatal diagnosis of human genetic disorders

XX Sequence 132 AA;

XX Query Match 100.0%; Score 50; DB 4; Length 132;

XX Best Local Similarity 100.0%; Pred. No. 3.8;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRPFRGR 8

Db 83 HRPFRGR 90

RESULT 7

AAAM72058

ID AAM72058 standard; protein; 132 AA.

XX AAM72058;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed probe encoded protein SEQ ID NO: 32364.

XX Human; bone marrow expressed exon; Gene expression analysis; probe; microarray; cancer; leukaemia; lymphoma; myeloma.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US0000668.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing gene expression in human bone marrow.

XX Example 4; SEQ ID NO 32364; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human bone marrow. They can be used to measure gene expression in bone marrow samples, which may enable the improved diagnosis and treatment of cancers such as lymphoma, leukaemia and myeloma. The present sequence is a protein encoded by one of the probes of the invention

XX Sequence 132 AA;

XX Query Match 100.0%; Score 50; DB 4; Length 132;

XX Best Local Similarity 100.0%; Pred. No. 3.8;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRPFRGR 8

Db 83 HRPFRGR 90

RESULT 8

AAAM59494

ID AAM59494 standard; protein; 132 AA.

XX AAM59494;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe encoded protein SEQ ID NO: 31599.

XX Human; brain expressed exon; Gene expression analysis; probe; microarray; Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.

XX Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US0000667.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human brains.

XX Example 4; SEQ ID NO 31599; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention

XX Sequence 132 AA;

XX Query Match 100.0%; Score 50; DB 4; Length 132;

XX Best Local Similarity 100.0%; Pred. No. 3.8;

XX Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRPFRGR 8

Db 83 HRPFRGR 90

RESULT 9

ABG53742

ID ABG53742 standard; peptide; 132 AA.

XX ABG53742;

XX 25-FEB-2003 (first entry)



DE Human liver peptide, SEQ ID NO 32390.

XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;  
KW hypercholesterolaemia; coronary heart disease.

XX Homo sapiens.

XX WO200157273-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000664.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488998/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing  
PT gene expression in human adult liver.

XX Claim 27; SEQ ID NO 32390; 658pp; English.

XX The invention relates to a single exon nucleic acid probe (SENP) (I) for  
CC measuring human gene expression in a sample derived from human adult  
CC liver, comprising one of 13109 defined nucleotide sequences given in the  
CC specification (or complements/ fragments). The probe hybridises at high  
CC stringency to a nucleic acid molecule expressed in the human adult liver.  
CC (I) may be used for predicting, measuring and displaying gene expression  
CC in samples derived from human adult liver. The genes identified may be  
CC involved in genetic liver diseases such as cirrhosis,  
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
CC associated with coronary heart disease. ABG47348-ABG59930 represent human  
CC liver single exon encoded peptides of the invention. Note: The sequence  
CC information for this patent does not appear in the printed specification  
CC but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 132 AA;

Query Match 100.0%; Score 50; DB 4; Length 132;  
Best Local Similarity 100.0%; Pred. No. 3.8;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HPRPPRGR 8

Db 83 HPRPPRGR 90

RESULT 10

ABG41873

ID ABG41873 standard; peptide; 132 AA.

XX AC ABG41873;

XX 19-AUG-2002 (first entry)

XX Human peptide encoded by genome-derived single exon probe SEQ ID 31538.

XX Human; single exon probe; asthma; lung cancer; COPD; ILD;  
KW chronic obstructive pulmonary disease; interstitial lung disease;  
KW familial idiopathic pulmonary fibrosis; neurofibromatosis;  
KW tuberosus sclerosis; Gaucher's disease; Niemann-Pick disease;  
KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;  
KW pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;  
KW primary ciliary dyskinesia; pulmonary hypertension;  
XX hyaline membrane disease.

XX Homo sapiens.

XX WO200186003-A2.

XX 15-NOV-2001.

XX 30-JAN-2001; 2001WO-US000665.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2002-114183/15.

XX Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples.

XX Claim 27; SEQ ID NO 31538; 634pp; English.

XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of probes  
CC; the novel set of probes which hybridise at high stringency to a nucleic  
CC acid expressed in the human lung; measuring gene expression in a sample  
CC derived from human lung, comprising (a) contacting the array with a  
CC collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of the  
CC array; identifying exons in a eukaryotic genome, comprising (a)  
CC algorithmically predicting at least one exon from genomic sequences of  
CC the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene expression  
CC analysis, and for identifying exons in a gene, particularly using human  
CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a peptide/protein encoded by a single exon probe of  
CC the invention. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 132 AA;

Query Match 100.0%; Score 50; DB 5; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 3.8;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRPFRGR 8  
 DB 83 HRPFRGR 90

RESULT 11  
 ABR57423  
 ID ABR57423 standard; protein; 149 AA.  
 XX  
 AC ABR57423;  
 XX  
 DT 15-SEP-2003 (first entry)  
 XX  
 DE Human NOV7 protein SEQ ID NO:24.  
 XX  
 KW Human; NOVX; cytostatic; cardiant; antiinflammatory; immunosuppressive;  
 KW antiatherogenic; haemostatic; anti-HIV; antidiabetic; antiarteriosclerotic;  
 KW anorectic; antiasthmatic; nephrotropic; antiarthritic; hepatotropic;  
 KW neuroprotective; nootropic; antibacterial; virucide; antiparasitic;  
 KW relaxant; anticonvulsant; hypotensive; vasotropic; antiparkinsonian;  
 KW vulnary; angiogenic; antiangiogenic; gene therapy; vaccine; cancer;  
 KW cardiomyopathy; atherosclerosis; hypertension; diabetes; inflammation;  
 KW autoimmune disorder; allergy; blood disorder; AIDS; obesity; asthma;  
 KW acquired immunodeficiency syndrome; nephropathy; cirrhosis; arthritis;  
 KW Alzheimer's disease; Parkinson's disease; Goitre; infection; stroke;  
 KW muscular dystrophy; epilepsy; wasting disorder; chromosome 12.  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO200294870-A2.  
 XX  
 PD 28-NOV-2002.  
 XX  
 PF 02-NOV-2001; 2001WO-US051580.  
 XX  
 PR 02-NOV-2000; 2000US-0245291P.  
 PR 02-NOV-2000; 2000US-0245317P.  
 PR 08-NOV-2000; 2000US-0246562P.  
 PR 26-JAN-2001; 2001US-0264389P.  
 PR 26-JAN-2001; 2001US-0264423P.  
 PR 29-JAN-2001; 2001US-0264799P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Groese WM, Macdougall JR, Smithson G, Millet I, Stone DJ;  
 PI Gunther E, Ellerman K, Alsobrook JP, Lepley DM, Burgess CE;  
 PI Spytek KA, Edinger SR, Gangoli EA, Gorman L, Taupier RJ, Li L;  
 PI Guo X, Fernandes ER, Vernet CAM, Tchernev VT, Casman SJ, Sheny S;  
 PI Mishra V, Furtak K, Baugartner JC, Colman SD;  
 XX  
 DR WPI: 2003-140359/13.  
 DR N-PSDB; ACF03558.  
 XX  
 PT New NOVX polypeptide useful for preventing or treating NOVX-associated  
 PT disorders, e.g. cancer, cardiomyopathy, atherosclerosis or diabetes, and  
 PT in chromosome mapping, tissue typing or pharmacogenomics.  
 XX  
 PS Claim 1; Page 69; 346pp; English.  
 XX  
 CC ACF03547 to ACF03570 encode the human NOVX proteins (I) given in ABR57412  
 CC to ABR57435. (I) have cytostatic, cardiant, antiinflammatory, nootropic,  
 CC immunosuppressive, antiallergic, haemostatic, anti-HIV, antidiabetic,  
 CC antiarteriosclerotic, anorectic, antiasthmatic, nephrotropic, virucide,  
 CC antiparasitic, hepatotropic, neuroprotective, antibacterial, relaxant,  
 CC antiparkinsonian, anticonvulsant, hypotensive, vasotropic, antiparkinsonian,  
 CC vulnary, angiogenic and antiangiogenic activities, and can be used in  
 CC gene therapy and vaccines. The NOVX polypeptides and their antibodies can  
 CC be used to determine the presence or absence of (I) in a sample. The NOVX

CC polypeptides, polynucleotides encoding them, and antibodies against them,  
 CC are useful in manufacturing a medicament for treating or preventing a  
 CC syndrome associated with a NOVX-associated disorder such as hypertension,  
 CC cardiomyopathy, atherosclerosis, cancer, diabetes, asthma, inflammation,  
 CC autoimmune disorders, allergies, blood disorders, obesity, acquired  
 CC immunodeficiency syndrome (AIDS), immunoglobulin (IgA) nephropathy,  
 CC cirrhosis, arthritis, Alzheimer's disease, Parkinson's disease, goitre,  
 CC infections (e.g. bacterial, viral, parasitic), stroke, muscular  
 CC dystrophy, epilepsy, and other wasting disorders associated with chronic  
 CC diseases. ACF03571 to ACF03644 represent PCR primers and probes for NOVX  
 CC sequence, which are used in an example from the present invention  
 XX  
 SQ Sequence 149 AA;

Query Match 100.0%; Score 50; DB 6; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 4.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRPFRGR 8  
 DB 100 HRPFRGR 107

RESULT 12  
 ABR56769  
 ID ABR56769 standard; protein; 154 AA.  
 XX  
 AC ABR56769;  
 XX  
 DT 30-JUL-2003 (first entry)  
 XX  
 DE Human secreted protein SSCP-44 SEQ ID NO:44.  
 XX  
 KW Human; secreted protein; SSCP; cytostatic; antiarteriosclerotic;  
 KW anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;  
 KW antiatherogenic; antiinflammatory; thyromimetic; gene therapy; cancer;  
 KW cell proliferative disorder; atherosclerosis; neurological disorder;  
 KW epilepsy; Huntington's disease; stroke; immune disorder; allergy; AIDS;  
 KW inflammatory disorder; developmental disorder; hypothyroidism;  
 XX  
 OS Homo sapiens.  
 XX  
 FN WO2003016506-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 15-AUG-2002; 2002WO-US027143.  
 XX  
 PR 17-AUG-2001; 2001US-0313249P.  
 PR 24-AUG-2001; 2001US-0314752P.  
 PR 07-SEP-2001; 2001US-0317818P.  
 PR 07-SEP-2001; 2001US-0317824P.  
 PR 21-SEP-2001; 2001US-0324040P.  
 PR 24-SEP-2001; 2001US-0324586P.  
 PR 02-NOV-2001; 2001US-0343980P.  
 PR 28-NOV-2001; 2001US-0343229P.  
 PR 13-FEB-2002; 2002US-0357002P.  
 PR 06-MAR-2002; 2002US-0362439P.  
 PR 19-MAR-2002; 2002US-0366041P.  
 PR 30-APR-2002; 2002US-0376988P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 PI Tang TY, Warren BA, Gietzen KJ, Lal PG, Yue H, Honchell CD;  
 PI Lehr-Mason PM, Burford N, Xu Y, Baughn MR, Duggan BM, Tran UK;  
 PI Lee EA, Forsythe IU, Richardson TW, Lee S, Thangavelu K, Yue H;  
 PI Emerling BM, Walla NK, Azimzal Y, Sanjanwala B, Hafalia AJA;  
 PI Borowsky ML, Nguyen DB, Ison CH, Astronomoff A, Ding L, Lee SY;  
 PI Becha SD, Rankumar J, Gardhi AR, Jin P, Fu GK, Swarnakar A;  
 XX  
 DR WPI: 2003-278569/27.  
 DR N-PSDB; ACC79069.

XX New human secreted proteins (SECP), useful for diagnosing, treating and  
PT preventing diseases or conditions associated with the aberrant SECP  
PT expression e.g. cancer, AIDS, atherosclerosis, Huntington's disease,  
XX stroke, infections.

XX Claim 1; Page 222; 286pp; English.

XX ACC79026 to ACC79105 encode the human secreted proteins (I) given in  
CC ABR56726 to ABR56805, designated SECP-1 to SECP-80. SECP sequences can  
CC have cytostatic, antiarteriosclerotic, anticonvulsant, antiinflammatory,  
CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic and  
CC thrombinetic activities, and can be used in gene therapy. The SECP  
CC proteins and polynucleotides can be used in diagnosing, treating and  
CC preventing diseases or conditions associated with the decreased  
CC expression or overexpression of SECP, such as cell proliferative (e.g.  
CC cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's  
CC disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and  
CC developmental (e.g. hypothyroidism, Cushing's syndrome) disorders, or  
CC infections. They are also useful in assessing the effects of exogenous  
CC compounds on the expression of nucleic acid and amino acid sequences of  
CC SECP. The SECP or its fragments are useful in screening compounds for  
CC effectiveness as agonist or antagonist of the polypeptides, or in  
CC altering the expression of the target polynucleotide and compounds that  
CC specifically bind to or modulate the activity of the polypeptide

XX Sequence 154 AA;

Query Match 100.0%; Score 50; DB 6; Length 154;  
Best Local Similarity 100.0%; Pred. No. 4.3;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRPFRPRGR 8  
|||||

Db 105 HRPFRPRGR 112

RESULT 13

ADA83798

ID ADA83798 standard; protein; 166 AA.

XX ADA83798;

AC ADA83798;

XX 20-NOV-2003 (first entry)

DT Human PRH2 protein.

DE human; marker; expressed sequence tag; EST; arabidopsis; tumour;  
XX stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;  
XX vaccine.

XX Homo sapiens.

OS WO2002103028-A2.

PN 27-DEC-2002.

PD 30-MAY-2002; 2002WO-IB004189.

PF 30-MAY-2001; 2001US-0293999P.

PR 22-OCT-2001; 2001US-0330457P.

PR 19-FEB-2002; 2002US-0357144P.

XX (BIOM-) BIOMEDICAL CENT.

XX Baranova AV, Yankovsky NK, Kozlov AP, Lobashev AV, Krukovskaya LL;  
XX WPI; 2003-175241/17.

DR N-PSDB; ADA83797.

DR Determining if a nucleic acid is a marker for a phenotype/cell type of  
XX interest, by global comparison of expressed sequence tags known to be  
XX expressed in the phenotype/cell type with all ESTs expressed in normal

PT tissue.

XX Claim 29; Page 191-192; 516pp; English.

PS The invention relates to a novel method for determining if a nucleic acid  
XX is a marker for a predetermined phenotype/cell type of interest from a  
XX biological species. The method comprises performing a global comparison  
XX of a group of expressed sequence tags (ESTs) known to be expressed in the  
XX phenotype/cell type of interest with all ESTs expressed in normal tissue  
XX in order to identify ESTs that are preferentially expressed in the  
XX phenotype/cell of interest. A method of the invention is useful for  
XX determining whether a nucleic acid is a marker for a predetermined  
XX phenotype or cell type of interest from a biological species, preferably  
XX Arabidopsis or human. The cell type of interest is an abnormal cell such  
XX as a tumour cell, and the predetermined phenotype is a stress-induced  
XX phenotype such as hyperosmotic stress or high salt conditions. A method  
XX of the invention is also useful for determining the progression of colon  
XX cancer in a human, for detecting a tumour cell, and for regulating or  
XX preventing the growth of a tumour cell. An antibody of the invention is  
XX useful for detecting the absence or presence of peptides encoded by  
XX tumour-associated markers. A polypeptide of the invention is useful as an  
XX immunogen for vaccinating an animal. The present sequence represents a  
XX tumour-associated antigen of the invention.

XX Sequence 166 AA;

Query Match 100.0%; Score 50; DB 6; Length 166;  
Best Local Similarity 100.0%; Pred. No. 4.6;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRPFRPRGR 8  
|||||

Db 117 HRPFRPRGR 124

RESULT 14

ADC98216

ID ADC98216 standard; protein; 166 AA.

XX ADC98216;

AC ADC98216;

XX 01-JAN-2004 (first entry)

DT Human salivary acidic proline-rich phosphoprotein (PRP).

DE Human; salivary acidic proline-rich phosphoprotein; PRP; autoantigen;  
XX atopic dermatitis-inducing protein; salivary gland; IgE autoantibody;  
XX immunoglobulin E; mast cell activation; basophil activation; diagnosis;  
XX risk assessment; sensitisation remedy; dermatological; antiallergic;  
XX antiinflammatory.

XX Homo sapiens.

OS WO2003084991-A1.

PN 16-OCT-2003.

PD 04-APR-2003; 2003WO-JP004325.

PF 08-APR-2002; 2002JP-00105425.

PR (NISC-) JAPAN SCI & TECHNOLOGY CORP.

PA Hide M, Yamamoto S, Tanaka T, Koro O;

PI WPI; 2003-833567/77.

XX N-PSDB; ADC98215.

DR Atopic dermatitis-inducing proteins, applicable in diagnosis of including  
XX risk of onset, and in developing sensitization remedies for the disease.

XX Claim 4; SEQ ID NO 2; 43pp; Japanese.

CC The invention relates to the human atopic dermatitis-inducing proteins,  
 CC salivary acidic proline-rich phosphoprotein (PRP; ADC98216) and prolactin  
 CC -inducible protein (PIP; ADC98218), and their post-translationally  
 CC modified forms. These proteins are secreted by salivary or sweat glands  
 CC and bind to IgE autoantibodies, thereby activating mast cells and  
 CC basophils. The invention also relates to antigenic peptide fragments of  
 CC PRP or PIP; antibodies which bind to PRP or PIP; methods for diagnosing  
 CC atopic dermatitis or for determining whether an individual is at risk of  
 CC developing atopic dermatitis by determining the presence of PRP- or PIP-  
 CC specific antibodies or immune complexes, or by quantifying histamine  
 CC release, and sensitisation remedies for atopic dermatitis containing PRP  
 CC and/or PIP or their peptides as the active ingredient. PRP, PIP and their  
 CC antibodies are useful in diagnosing atopic dermatitis, or for determining  
 CC whether an individual is at risk of developing atopic dermatitis. They  
 CC are also useful in developing sensitisation remedies for the treatment of  
 CC atopic dermatitis. The present sequence represents the specifically  
 CC claimed human salivary acidic proline-rich phosphoprotein (PRP).

XX Sequence 166 AA;

Query Match 100.0%; Score 50; DB 7; Length 166;

Best Local Similarity 100.0%; Pred. No. 4.6; 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Gaps 0;

QY 1 HPRPPRGR 8

DB 117 HPRPPRGR 124

RESULT 15

AAG58448

ID AAG58448 standard; protein; 51 AA.

AC AAG58448;

XX

XX

DT 18-OCT-2000 (first entry)

XX

DE Zea mays protein fragment SEQ ID NO: 75449.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence; corn.

XX Zea mays subsp. mays.

XX EF1033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-00301439.

XX 25-FEB-1999; 99US-0121825P.

XX 05-MAR-1999; 99US-0123180P.

XX 09-MAR-1999; 99US-0123548P.

XX 23-MAR-1999; 99US-0125788P.

XX 29-MAR-1999; 99US-0126284P.

XX 01-APR-1999; 99US-0126785P.

XX 08-APR-1999; 99US-0127462P.

XX 16-APR-1999; 99US-0128234P.

XX 19-APR-1999; 99US-0128714P.

XX 21-APR-1999; 99US-0130077P.

XX 23-APR-1999; 99US-0130449P.

XX 28-APR-1999; 99US-0130510P.

XX 30-APR-1999; 99US-0130891P.

XX 04-MAY-1999; 99US-0131449P.

XX 05-MAY-1999; 99US-0132048P.

XX 06-MAY-1999; 99US-0132407P.

XX 06-MAY-1999; 99US-0132484P.

XX 06-MAY-1999; 99US-0132485P.

XX 06-MAY-1999; 99US-0132486P.

XX 06-MAY-1999; 99US-0132487P.

XX 07-MAY-1999; 99US-0132863P.

PR 11-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 14-MAY-1999; 99US-0134370P.

PR 14-MAY-1999; 99US-0134768P.

PR 18-MAY-1999; 99US-0134941P.

PR 19-MAY-1999; 99US-0135124P.

PR 21-MAY-1999; 99US-0135353P.

PR 24-MAY-1999; 99US-0135629P.

PR 25-MAY-1999; 99US-0136021P.

PR 27-MAY-1999; 99US-0136392P.

PR 28-MAY-1999; 99US-0136782P.

PR 01-JUN-1999; 99US-0137222P.

PR 03-JUN-1999; 99US-0137528P.

PR 04-JUN-1999; 99US-0137502P.

PR 07-JUN-1999; 99US-0137724P.

PR 08-JUN-1999; 99US-0138094P.

PR 10-JUN-1999; 99US-0138540P.

PR 10-JUN-1999; 99US-0138847P.

PR 14-JUN-1999; 99US-0139119P.

PR 16-JUN-1999; 99US-0139452P.

PR 17-JUN-1999; 99US-0139453P.

PR 18-JUN-1999; 99US-0139492P.

PR 18-JUN-1999; 99US-0139454P.

PR 18-JUN-1999; 99US-0139455P.

PR 18-JUN-1999; 99US-0139456P.

PR 18-JUN-1999; 99US-0139457P.

PR 18-JUN-1999; 99US-0139458P.

PR 18-JUN-1999; 99US-0139459P.

PR 18-JUN-1999; 99US-0139460P.

PR 18-JUN-1999; 99US-0139461P.

PR 18-JUN-1999; 99US-0139462P.

PR 18-JUN-1999; 99US-0139463P.

PR 18-JUN-1999; 99US-0139750P.

PR 18-JUN-1999; 99US-0139763P.

PR 21-JUN-1999; 99US-0139817P.

PR 22-JUN-1999; 99US-0139899P.

PR 23-JUN-1999; 99US-0140353P.

PR 23-JUN-1999; 99US-0140354P.

PR 24-JUN-1999; 99US-0140695P.

PR 28-JUN-1999; 99US-0140823P.

PR 29-JUN-1999; 99US-0140991P.

PR 30-JUN-1999; 99US-0141287P.

PR 01-JUL-1999; 99US-0141542P.

PR 01-JUL-1999; 99US-0142154P.

PR 02-JUL-1999; 99US-0142055P.

PR 06-JUL-1999; 99US-0142390P.

PR 08-JUL-1999; 99US-0142803P.

PR 09-JUL-1999; 99US-0142920P.

PR 12-JUL-1999; 99US-0142977P.

PR 13-JUL-1999; 99US-0143542P.

PR 14-JUL-1999; 99US-0143624P.

PR 15-JUL-1999; 99US-0144005P.

PR 16-JUL-1999; 99US-0144086P.

PR 16-JUL-1999; 99US-0144086P.

PR 19-JUL-1999; 99US-0144325P.

PR 19-JUL-1999; 99US-0144331P.

PR 19-JUL-1999; 99US-0144332P.

PR 19-JUL-1999; 99US-0144333P.

PR 19-JUL-1999; 99US-0144334P.

PR 19-JUL-1999; 99US-0144335P.

PR 19-JUL-1999; 99US-0144352P.

PR 20-JUL-1999; 99US-0144632P.

PR 20-JUL-1999; 99US-0144884P.

PR 20-JUL-1999; 99US-0144814P.

PR 21-JUL-1999; 99US-0145086P.

PR 21-JUL-1999; 99US-0145088P.

PR 22-JUL-1999; 99US-0145089P.

PR 22-JUL-1999; 99US-0145087P.

PR 22-JUL-1999; 99US-0145089P.

PR 22-JUL-1999; 99US-0145192P.

PR 23-JUL-1999; 99US-0145145P.

PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 21-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155133P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 05-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 14-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.

PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Query Match 88.0%; Score 44; DB 3; Length 51;  
Best Local Similarity 87.5%; Pred. No. 13;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HRPFRGR 8  
|||  
Db 23 HRPFRSR 30

Search completed: April 6, 2004, 16:06:40  
Job time : 45.785 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 16:14:50 ; Search time 31.4019 Seconds  
(without alignments)  
66.909 Million cell updates/sec

Title: US-10-009-709-9

Perfect score: 50

Sequence: 1 HRRPRGR 8

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 26263353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	132	9	US-09-864-761-43644
2	50	100.0	166	14	US-10-157-031-80
3	45	90.0	238	12	US-10-425-114-69167
4	45	90.0	277	12	US-10-425-114-43478
5	42	84.0	72	12	US-10-424-599-223703
6	42	84.0	19652	15	US-10-084-846A-7
7	41	82.0	2796	9	US-09-870-759-114
8	41	82.0	2796	10	US-09-751-708A-114
9	39	78.0	71	12	US-10-424-599-17811
10	39	78.0	164	12	US-10-424-599-17830
11	39	78.0	198	12	US-10-424-599-239718
12	39	78.0	278	12	US-10-425-114-66403
13	39	78.0	470	12	US-10-282-122A-62669
14	39	78.0	618	14	US-10-406-209-5
15	38	76.0	66	14	US-10-029-386-29711

16	38	76.0	94	12	US-10-424-599-285441
17	38	76.0	151	14	US-10-106-698-6771
18	38	76.0	276	12	US-10-425-114-71744
19	37	74.0	40	14	US-10-008-524A-106
20	37	74.0	40	15	US-10-350-719-106
21	37	74.0	46	9	US-09-864-761-38759
22	37	74.0	71	12	US-10-424-599-151579
23	37	74.0	95	9	US-09-925-300-1824
24	37	74.0	96	12	US-10-424-599-191513
25	37	74.0	131	12	US-10-425-114-52873
26	37	74.0	165	12	US-10-424-599-245792
27	37	74.0	196	12	US-10-425-114-69376
28	37	74.0	206	14	US-10-156-761-14830
29	37	74.0	209	15	US-10-457-372-7
30	37	74.0	214	15	US-10-104-047-2741
31	37	74.0	219	16	US-10-389-566-1429
32	37	74.0	245	8	US-08-424-550B-38
33	37	74.0	247	15	US-10-108-260A-2734
34	37	74.0	262	12	US-10-425-114-59553
35	37	74.0	276	12	US-10-424-599-279857
36	37	74.0	291	12	US-10-425-114-60421
37	37	74.0	401	15	US-10-369-493-5026
38	37	74.0	429	9	US-09-827-822-6
39	37	74.0	440	14	US-10-157-031-66
40	37	74.0	474	12	US-10-425-114-54425
41	37	74.0	513	12	US-10-425-114-52040
42	37	74.0	654	12	US-10-425-114-51968
43	37	74.0	674	11	US-09-981-151A-70
44	37	74.0	679	11	US-09-981-151A-69
45	37	74.0	687	10	US-09-934-455-18

#### ALIGNMENTS

#### RESULT 1

US-09-864-761-43644  
; Sequence 43644, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aemica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00663  
; PRIOR FILING DATE: 2001-01-30

Sequence 285441,  
Sequence 6771, Ap  
Sequence 71744, A  
Sequence 106, App  
Sequence 106, App  
Sequence 38759, A  
Sequence 151579,  
Sequence 1824, Ap  
Sequence 191513,  
Sequence 52873, A  
Sequence 245792,  
Sequence 69576, A  
Sequence 14830, A  
GENERAL INFORMA  
Sequence 2741, Ap  
Sequence 1429, Ap  
Sequence 38, Appl  
Sequence 2734, Ap  
Sequence 59553, A  
Sequence 279857, A  
Sequence 60421, A  
Sequence 5026, Ap  
Sequence 6, Appli  
Sequence 54425, A  
Sequence 52040, A  
Sequence 51968, A  
Sequence 70, Appl  
Sequence 69, Appl  
Sequence 18, Appl

; PRIOR APPLICATION NUMBER: PCT/US01/00662  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00661  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00670  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: US 60/234,687  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: US 09/608,408  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: US 09/774,203  
; PRIOR FILING DATE: 2001-03-29  
; NUMBER OF SEQ ID NOS: 49117  
; SOFTWARE: Annotmax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 43644  
; LENGTH: 132  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO AC006518.17  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77  
; OTHER INFORMATION: EST HUMAN HIT: BF088785.1, EVALUE 1.00e-01  
; OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02  
US-09-864-761-43644

Query Match 100.0%; Score 50; DB 9; Length 132;  
Best Local Similarity 100.0%; Pred.No. 6.9;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRPFRGR 8  
Db 83 HRPFRGR 90

## RESULT 2

US-10-157-031-80  
; Sequence 80, Application US/10157031  
; Publication No. US20030108890A1  
; GENERAL INFORMATION:  
; APPLICANT: Baranova, A. V.  
; APPLICANT: Yankovsky, N. K.  
; APPLICANT: Kozlov, A. P.  
; APPLICANT: Lobashev, A. V.  
; APPLICANT: Kravskaya, L. L.  
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences  
; FILE REFERENCE: 2760-103  
; CURRENT APPLICATION NUMBER: US/10/157,031  
; CURRENT FILING DATE: 2002-05-30  
; NUMBER OF SEQ ID NOS: 415  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 80  
; LENGTH: 166  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-157-031-80

Query Match 100.0%; Score 50; DB 14; Length 166;  
Best Local Similarity 100.0%; Pred.No. 8.4;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRPFRGR 8  
Db 117 HRPFRGR 124

## RESULT 3

US-10-425-114-69167  
; Sequence 69167, Application US/10425114

; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 69167  
; LENGTH: 238  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17307C10\_FLI.pep  
US-10-425-114-69167

Query Match 90.0%; Score 45; DB 12; Length 238;  
Best Local Similarity 100.0%; Pred.No. 57;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRPFRGR 7  
Db 155 HRPFRGR 161

## RESULT 4

US-10-425-114-43478  
; Sequence 43478, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 43478  
; LENGTH: 277  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: 700423863\_FLI.pep  
US-10-425-114-43478

Query Match 90.0%; Score 45; DB 12; Length 277;  
Best Local Similarity 100.0%; Pred.No. 65;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRPFRGR 7  
Db 86 HRPFRGR 92

## RESULT 5

US-10-424-599-223703  
; Sequence 223703, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua

```

; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 223703
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_44033C.1.pep
US-10-424-599-223703

Query Match      84.0%; Score 42; DB 12; Length 72;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PRPPRGR 8
Db 27 PRPPRGR 33

RESULT 6
US-10-084-846A-7
; Sequence 7, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patentin Ver. 3.2
; SEQ ID NO 7
; LENGTH: 19652
; TYPE: PRT
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 2.
; OTHER INFORMATION: Start codon: gat, Start position: nucleotide 2.
US-10-084-846A-7

Query Match      84.0%; Score 42; DB 15; Length 19652;
Best Local Similarity 100.0%; Pred. No. 6.6e+03;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PRPPRGR 8
Db 2260 PRPPRGR 2266

RESULT 7
US-09-870-759-114
; Sequence 114, Application US/09870759
; Patent No. US2002017551A1
; GENERAL INFORMATION:
; APPLICANT: TERMAN, David S
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
; FILE REFERENCE: 870759
; CURRENT APPLICATION NUMBER: US/09/870,759
; CURRENT FILING DATE: 2002-01-14
; PRIOR APPLICATION NUMBER: US 60/208,128
; PRIOR FILING DATE: 2000-05-30

```

```

; NUMBER OF SEQ ID NOS: 166
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 114
; LENGTH: 2796
; TYPE: PRT
; ORGANISM: Mycobacterium bovis
US-09-870-759-114

Query Match      82.0%; Score 41; DB 9; Length 2796;
Best Local Similarity 87.5%; Pred. No. 1.7e+03;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HRPFRGR 8
Db 422 HRPFRGR 429

RESULT 8
US-09-751-708A-114
; Sequence 114, Application US/09751708A
; Publication No. US20030157113A1
; GENERAL INFORMATION:
; APPLICANT: TERMAN, David S
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT OF NEOPLASTIC DISEASE
; FILE REFERENCE: 751708
; CURRENT APPLICATION NUMBER: US/09/751,708A
; CURRENT FILING DATE: 2002-10-15
; PRIOR APPLICATION NUMBER: US 60/173,371
; PRIOR FILING DATE: 1999-12-28
; NUMBER OF SEQ ID NOS: 166
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 114
; LENGTH: 2796
; TYPE: PRT
; ORGANISM: Mycobacterium bovis
US-09-751-708A-114

Query Match      82.0%; Score 41; DB 10; Length 2796;
Best Local Similarity 87.5%; Pred. No. 1.7e+03;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HRPFRGR 8
Db 422 HRPFRGR 429

RESULT 9
US-10-424-599-178111
; Sequence 178111, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 178111
; LENGTH: 71
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_131850C.1.pep
US-10-424-599-178111

Query Match      78.0%; Score 39; DB 12; Length 71;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```



Qy 2 PRPPRGR 8  
Db 61 PRPPRGR 67

RESULT 10  
US-10-424-599-174830  
; Sequence 174830, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 174830  
; LENGTH: 164  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_128891C.1.pap  
US-10-424-599-174830

Query Match 78.0%; Score 39; DB 12; Length 164;  
Best Local Similarity 75.0%; Pred. No. 2.9e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 HRPGRGR 8  
Db 53 HEKPRGR 60

RESULT 11  
US-10-424-599-239718  
; Sequence 239718, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 239718  
; LENGTH: 198  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(198)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_58490C.1.pap  
US-10-424-599-239718

Query Match 78.0%; Score 39; DB 12; Length 198;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRPGRGR 6  
Db 65 HRPGRGR 70

RESULT 12  
US-10-425-114-66403  
; Sequence 66403, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 66403  
; LENGTH: 278  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB4729-084-D3\_FLI.pap  
US-10-425-114-66403

Query Match 78.0%; Score 39; DB 12; Length 278;  
Best Local Similarity 100.0%; Pred. No. 4.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 HRPGRGR 6  
Db 25 HRPGRGR 30

RESULT 13  
US-10-282-122A-62669  
; Sequence 62669, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/230,347  
; PRIOR FILING DATE: 2000-09-09  
; PRIOR APPLICATION NUMBER: 60/242,578  
; PRIOR FILING DATE: 2000-10-23  
; PRIOR APPLICATION NUMBER: 60/253,625  
; PRIOR FILING DATE: 2000-11-27  
; PRIOR APPLICATION NUMBER: 60/257,931  
; PRIOR FILING DATE: 2000-12-22  
; PRIOR APPLICATION NUMBER: 60/267,636  
; PRIOR FILING DATE: 2001-02-09

; PRIOR APPLICATION NUMBER: 60/269,308  
 ; PRIOR FILING DATE: 2001-02-16  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 78614  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 62869  
 ; LENGTH: 470  
 ; TYPE: PRT  
 ; ORGANISM: Mycobacterium bovis  
 US-10-282-122A-62669

Query Match 78.0%; Score 39; DB 12; Length 470;  
 Best Local Similarity 85.7%; Pred. No. 7.2e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HRPFRPG 7  
 Db 72 HRPFRPG 78

RESULT 14  
 US-10-406-209-5  
 ; Sequence 5, Application US/10406209  
 ; Publication No. US20030170758A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KYOWA HAKKO KOGYO CO., LTD.  
 ; TITLE OF INVENTION: NOVEL ANTIBODIES, DRUGS CONTAINING THESE ANTIBODIES AND  
 ; TITLE OF INVENTION: METHODS FOR  
 ; TITLE OF INVENTION: SCREENING COMPOUNDS BY USING THESE ANTIBODIES  
 ; FILE REFERENCE: 1241.19  
 ; CURRENT APPLICATION NUMBER: US/10/406,209  
 ; CURRENT FILING DATE: 2003-04-04  
 ; PRIOR APPLICATION NUMBER: US/09/806,228C  
 ; PRIOR FILING DATE: 2001-08-30  
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: PCT/JP99/05350  
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1999-09-29  
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: JP 10-291501  
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-29  
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: JP 10-291503  
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-09-29  
 ; NUMBER OF SEQ ID NOS: 28  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 5  
 ; LENGTH: 618  
 ; TYPE: PRT  
 ; ORGANISM: Mouse  
 US-10-406-209-5

Query Match 78.0%; Score 39; DB 14; Length 618;  
 Best Local Similarity 100.0%; Pred. No. 9.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HRPFRPG 6  
 Db 331 HRPFRPG 336

RESULT 15  
 US-10-029-386-29711  
 ; Sequence 29711, Application US/10029386  
 ; Publication No. US20030194704A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Penn, Sharon G.  
 ; APPLICANT: Rank, David R.  
 ; APPLICANT: Hanzel, David K.  
 ; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR C  
 ; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
 ; FILE REFERENCE: AEOMICA-X-2  
 ; CURRENT APPLICATION NUMBER: US/10/029,386  
 ; CURRENT FILING DATE: 2001-12-20  
 ; NUMBER OF SEQ ID NOS: 34288  
 ; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
 ; SEQ ID NO 29711

; LENGTH: 66  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: MAP TO CHR11.1  
 ; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3  
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.9  
 ; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8  
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.9  
 ; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4  
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.5  
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2  
 ; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3  
 ; OTHER INFORMATION: SWISSPROT HIT: P03291, EVALUE 3.50e+00  
 US-10-029-386-29711

Query Match 76.0%; Score 38; DB 14; Length 66;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HRPFRPG 7  
 Db 5 HRPFRPG 11

Search completed: April 6, 2004, 17:06:07  
 Job time : 32.4019 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:56:34 ; Search time 11.7383 Seconds  
(without alignments)  
35.185 Million cell updates/sec

Title: US-10-009-709-9

Perfect score: 50

Sequence: 1 HRPFRGR 8

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/6CTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/6CTUS\_COMB1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	50	100.0	418	4	US-09-252-991A-29452
2	42	84.0	207	4	US-09-252-991A-29505
3	42	84.0	223	4	US-09-252-991A-17072
4	42	84.0	259	4	US-09-252-991A-24480
5	41	82.0	152	4	US-09-252-991A-31405
6	41	82.0	439	4	US-09-252-991A-16736
7	41	82.0	676	4	US-09-252-991A-24713
8	39	78.0	225	4	US-09-252-991A-23933
9	39	78.0	302	4	US-09-252-991A-31918
10	39	78.0	420	4	US-09-252-991A-30229
11	39	78.0	422	4	US-09-252-991A-27513
12	39	78.0	594	4	US-09-252-991A-32578
13	39	78.0	611	4	US-09-252-991A-28230
14	38.5	77.0	471	4	US-09-252-991A-31856
15	38	76.0	259	4	US-09-252-991A-30663
16	38	76.0	377	4	US-09-252-991A-29332
17	38	76.0	453	4	US-09-252-991A-27041
18	38	76.0	638	4	US-09-252-991A-27068
19	38	76.0	845	4	US-09-252-991A-17856
20	37	74.0	40	4	US-09-314-268-106
21	37	74.0	155	4	US-09-252-991A-20099
22	37	74.0	182	4	US-09-252-991A-21772
23	37	74.0	208	4	US-09-252-991A-32430
24	37	74.0	209	4	US-08-778-717-7
25	37	74.0	226	3	US-09-195-868-30
26	37	74.0	245	4	US-08-463-260A-38
27	37	74.0	245	4	US-08-488-446-38

28 37 74.0 245 4 US-08-467-344A-38 Sequence 38, Appl  
29 37 74.0 256 4 US-09-252-991A-19839 Sequence 19839, A  
30 37 74.0 266 4 US-09-252-991A-21719 Sequence 21719, A  
31 37 74.0 355 3 US-08-483-533-41 Sequence 41, Appl  
32 37 74.0 355 4 US-09-283-471A-41 Sequence 41, Appl  
33 37 74.0 355 5 PCT-US91-06532-3 Sequence 3, Appl  
34 37 74.0 429 1 US-07-672-883-3 Sequence 3, Appl  
35 37 74.0 429 3 US-09-370-368-6 Sequence 6, Appl  
36 37 74.0 429 6 5182195-13 Patent No. 5182195  
37 37 74.0 474 3 US-08-729-416C-1 Sequence 1, Appl  
38 37 74.0 474 3 US-08-729-416C-11 Sequence 11, Appl  
39 37 74.0 474 3 US-08-807-342B-2 Sequence 2, Appl  
40 37 74.0 474 4 US-09-433-353-1 Sequence 1, Appl  
41 37 74.0 474 4 US-09-433-353-11 Sequence 11, Appl  
42 37 74.0 492 4 US-09-252-991A-30857 Sequence 30857, A  
43 37 74.0 585 4 US-09-252-991A-31529 Sequence 31529, A  
44 37 74.0 594 3 US-08-729-416C-7 Sequence 7, Appl  
45 37 74.0 594 3 US-08-729-416C-17 Sequence 17, Appl

## ALIGNMENTS

### RESULT 1

US-09-252-991A-29452  
; Sequence 29452, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 29452  
; LENGTH: 418  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29452

Query Match 100.0%; Score 50; DB 4; Length 418;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 HRPFRGR 8

Db 160 HRPFRGR 167

### RESULT 2

US-09-252-991A-29505  
; Sequence 29505, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 29505  
; LENGTH: 207  
; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29505

Query Match 84.0%; Score 42; DB 4; Length 207;  
Best Local Similarity 100.0%; Pred. No. 17;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PRPPRGR 8  
Db 84 PRPPRGR 90

## RESULT 3

US-09-252-991A-17072  
Sequence 17072, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252.991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 17072

; LENGTH: 223

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-17072

Query Match 84.0%; Score 42; DB 4; Length 223;

Best Local Similarity 87.5%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 1; Indels 1; Gaps 0;

QY 1 HRPFRGR 8

Db 139 HRPFRGR 146

## RESULT 4

US-09-252-991A-24480

; Sequence 24480, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252.991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 24480

; LENGTH: 559

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-24480

Query Match 84.0%; Score 42; DB 4; Length 559;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 PRPPRGR 8

Db 315 PRPPRGR 321

## RESULT 5

US-09-252-991A-31405

; Sequence 31405, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252.991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 31405

; LENGTH: 152

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-31405

Query Match 82.0%; Score 41; DB 4; Length 152;

Best Local Similarity 87.5%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HRPFRGR 8

Db 79 HRPFRGR 86

## RESULT 6

US-09-252-991A-16736

; Sequence 16736, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252.991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 16736

; LENGTH: 439

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-16736

Query Match 82.0%; Score 41; DB 4; Length 439;

Best Local Similarity 87.5%; Pred. No. 48;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HRPFRGR 8

Db 117 HRPFRGR 124

## RESULT 7

US-09-252-991A-24713

; Sequence 24713, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252.991A

; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24713  
; LENGTH: 676  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24713

Query Match 82.0%; Score 41; DB 4; Length 676;  
Best Local Similarity 87.5%; Pred. No. 72;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HRPFRGR 8  
Db 616 HRPFRGR 623

RESULT 8  
US-09-252-991A-23933  
; Sequence 23933, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23933  
; LENGTH: 225  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23933

Query Match 78.0%; Score 39; DB 4; Length 225;  
Best Local Similarity 85.7%; Pred. No. 51;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HRPFRGR 7  
Db 88 HRPFRGR 94

RESULT 9  
US-09-252-991A-31918  
; Sequence 31918, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 31918  
; LENGTH: 302  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-31918

Query Match 78.0%; Score 39; DB 4; Length 302;  
Best Local Similarity 85.7%; Pred. No. 68;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HRPFRGR 7  
Db 212 HRPFRGR 218

RESULT 10  
US-09-252-991A-30229  
; Sequence 30229, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30229  
; LENGTH: 420  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30229

Query Match 78.0%; Score 39; DB 4; Length 420;  
Best Local Similarity 85.7%; Pred. No. 92;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 HRPFRGR 7  
Db 71 HRPFRGR 77

RESULT 11  
US-09-252-991A-27513  
; Sequence 27513, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27513  
; LENGTH: 422  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27513

Query Match 78.0%; Score 39; DB 4; Length 422;  
Best Local Similarity 75.0%; Pred. No. 93;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 HRPFRGR 8  
Db 376 HRPFRGR 383

RESULT 12

```
US-09-252-991A-32578
; Sequence 32578, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32578
; LENGTH: 594
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32578

Query Match      78.0%; Score 39; DB 4; Length 594;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 HRPFR 6
Db      194 HRPFR 199

RESULT 13
US-09-252-991A-28230
; Sequence 28230, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 28230
; LENGTH: 611
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28230

Query Match      78.0%; Score 39; DB 4; Length 611;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 HRPFR 6
Db      499 HRPFR 504

RESULT 14
US-09-252-991A-31856
; Sequence 31856, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: 107196.136
```

```
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31856
; LENGTH: 471
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31856

Query Match      77.0%; Score 38.5; DB 4; Length 471;
Best Local Similarity 72.7%; Pred. No. 1.2e+02;
Matches 8; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY      1 HRPFR 8
Db      413 HRPFR 423

RESULT 15
US-09-252-991A-30663
; Sequence 30663, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30663
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30663

Query Match      76.0%; Score 38; DB 4; Length 259;
Best Local Similarity 85.7%; Pred. No. 83;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 HRPFR 7
Db      240 HRPFR 246

Search completed: April 6, 2004, 16:19:41
Job time : 12.7383 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:52:34 ; Search time 7.85047 Seconds  
(without alignments)  
85.771 Million cell updates/sec

Title: US-10-009-709-10

Perfect score: 42

Sequence: 1 PRPGR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: PIR\_78.\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	166	1 PHUSC	salivary proline-r
2	42	100.0	166	2 B25372	salivary proline-r
3	42	100.0	171	2 A27307	proline-rich phosph
4	42	100.0	561	2 E70610	hypothetical prote
5	39	92.9	200	2 S55609	hypothetical prote
6	38	90.5	450	2 A83238	nitrotriacetate
7	37	88.1	123	2 D72579	hypothetical prote
8	37	88.1	137	2 A88637	protein W09G1.9 (
9	37	88.1	151	2 H82546	hypothetical prote
10	37	88.1	237	2 D40595	hypothetical prote
11	37	88.1	340	2 A24036	erythromycin resis
12	37	88.1	401	2 T32737	hypothetical prote
13	37	88.1	429	1 FOLJGH	probable gag poly
14	37	88.1	429	1 FOLJCN	gag polyprotein -
15	37	88.1	429	2 S06073	gag polyprotein -
16	37	88.1	438	2 B72654	probable histidyl-
17	37	88.1	513	2 B74844	anthranilate synth
18	37	88.1	1385	2 H88569	protein K03H1.5 [1
19	37	88.1	1409	2 S41038	hypothetical prote
20	37	88.1	1466	2 T17138	CLIAA protein - ra
21	37	88.1	1467	2 T18411	latrophilin-1, bra
22	37	88.1	1471	2 T17149	CLIBA protein - ra
23	37	88.1	1472	2 T18413	latrophilin-1, bra
24	37	88.1	1510	2 T17145	CLIAB protein - ra
25	37	88.1	1515	2 T17156	CLIBB protein - ra
26	36	85.7	302	2 H96811	protein F319.20 [1
27	36	85.7	381	2 S48049	cholecystokinin B
28	36	85.7	447	2 A47430	gastrin/cholecysto
29	36	85.7	450	2 JQ1614	gastrin receptor -

#### ALIGNMENTS

##### RESULT 1

###### PIHUSC

salivary proline-rich phosphoprotein precursor PRH2 [validated] - human

N;Alternate names: salivary acidic proline-rich protein PRH2

N;Contains: peptide P-C (basic proline-rich peptide IB-8b); proline-rich phosphoprotei

C;Species: Homo sapiens (man)

C;Date: 31-Mar-1981 #sequence revision 12-Apr-1996 #text change 08-Dec-2000

C;Accession: A25372; A19803; B57868; A92277; A92254; A94425; A91954; S02564; S02563; J

R;Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.

J. Biol. Chem. 260, 11123-11130, 1985

A;Title: Differential RNA splicing and post-translational cleavages in the human saliv

A;Reference number: A92492; MUID:85289325; PMID:2993301

A;Accession: A25372

A;Molecule type: mRNA

A;Residues: 1-166 <MA>

A;Cross-references: GB:K03202; NID:G130481; PIDN:AAA60183.1; PID:G190482

R;Schlesinger, D.H.; Hay, D.I.

Int. J. Pept. Protein Res. 17, 34-41, 1981

A;Title: Primary structure of the active tryptic fragments of human and monkey salivar

A;Reference number: A31757; MUID:81191179; PMID:7228490

A;Accession: A19803

A;Molecule type: protein

A;Residues: 17-46 <SCH>

R;Kim, H.S.; Maeda, N.

J. Biol. Chem. 261, 6712-6718, 1986

A;Title: Structures of two haellii-type genes in the human salivary proline-rich protei

A;Reference number: A57868; MUID:86196106; PMID:3009472

A;Accession: B57868

A;Molecule type: DNA

A;Residues: 1-166 <KW>

A;Cross-references: GB:M13058; NID:G190513; PIDN:AAA98808.1; PID:G190514

R;Wong, R.S.C.; Bennick, A.

J. Biol. Chem. 255, 5943-5948, 1980

A;Title: The primary structure of a salivary calcium-binding proline-rich phosphoprote

A;Reference number: A92277; MUID:80204369; PMID:7380845

A;Contents: protein C

A;Accession: A92277

A;Molecule type: protein

A;Residues: 17-19, N', 21-166 <WON>

A;Note: the amino-terminal 46 residues are involved with inhibiting hydroxyapatite for

R;Wong, R.S.C.; Hofmann, T.; Bennick, A.

J. Biol. Chem. 254, 4800-4808, 1979

A;Title: The complete primary structure of a proline-rich phosphoprotein from human sa

A;Reference number: A92254; MUID:79173237; PMID:438215

A;Contents: protein A

A;Accession: A92254

A;Molecule type: protein

A;Residues: 17-19, N', 21-122 <WO2>

R;Schlesinger, D.H.; Hay, D.I.

In Peptides: Structure and Biological Function (Proc. Sixth Amer. Peptide Symp.), Gros

A;Title: Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent

A;Reference number: A94425

30 36 85.7 452 2 A46195  
31 36 85.7 452 2 JC2459  
32 36 85.7 453 2 S32817  
33 36 85.7 519 2 T07026  
34 36 85.7 635 2 T07794  
35 36 85.7 840 2 T36829  
36 36 85.7 1003 2 T13856  
37 36 85.7 1820 2 A55494  
38 36 85.7 1974 2 T30010  
39 35 83.3 264 2 S75053  
40 35 83.3 352 2 G83636  
41 35 83.3 433 2 D84335  
42 35 83.3 466 1 P2WLR1  
43 35 83.3 550 2 A46419  
44 35 83.3 633 2 T02673  
45 35 83.3 784 2 AG2736

cholecystokinin B  
gastrin/cholecysto  
gastrin receptor -  
ethylene receptor  
ethylene receptor  
probable phenylala  
ker protein - frul  
latent transformin  
hypothetical prote  
hypothetical prote  
conserved hypothet  
hypothetical prote  
L2 protein - rhesu  
trophoblast-endoth  
heterogeneous nucl  
exoribonuclease [1

A;Accession: A94425  
 A;Molecule type: protein  
 A;Residues: 17-122 <SC2>  
 A;Note: the authors call this protein PRP-4  
 R;Isemura, S.; Saitoh, E.; Sanada, K.  
 J. Biochem. 87, 1071-1077, 1980  
 A;Title: The amino acid sequence of a salivary proline-rich peptide, P-C, and its relationship to the human salivary proline-rich proteins (PRP)  
 A;Reference number: A91954; MUID:80227634; PMID:7390979  
 A;Contents: Peptide P-C  
 A;Accession: A91954  
 A;Molecule type: protein  
 A;Residues: 123-166 <SE>  
 R;Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck  
 Biochem. J. 255, 15-21, 1988  
 A;Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-1 to PRP-6)  
 A;Reference number: S02562; MUID:89061650; PMID:3196309  
 A;Accession: S02564  
 A;Molecule type: protein  
 A;Residues: 17-166 <HAY>  
 A;Accession: S02563  
 A;Molecule type: protein  
 A;Residues: 47-71 <HAY>  
 R;Schlesinger, D.H.; Hay, D.I.  
 Int. J. Pept. Protein Res. 27, 373-379, 1986  
 A;Title: Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of the human salivary proline-rich phosphoprotein, PRP-1  
 A;Reference number: JP0106; MUID:86222916; PMID:3710693  
 A;Accession: JP0106  
 A;Molecule type: protein  
 A;Residues: 17-161, Q', 163-166 <SC3>  
 A;Experimental source: parotid gland  
 R;Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A;Title: Basic proline-rich proteins from human parotid saliva: relationships of the covalent structure to the primary structure  
 A;Reference number: A39355; MUID:91190884; PMID:1849422  
 A;Accession: G39355  
 A;Status: preliminary  
 A;Molecule type: protein  
 A;Residues: 123-166 <KAU>  
 R;Robinson, R.; Kaufman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A;Title: Primary structure and possible origin of the non-glycosylated basic proline-rich protein from human parotid saliva  
 A;Reference number: S06153; MUID:90088384; PMID:2688632  
 A;Accession: S06153  
 A;Molecule type: protein  
 A;Residues: 123-166 <ROB>  
 C;Comment: The proposed biological functions are a highly potent inhibitor of crystal growth  
 C;Genetics:  
 A;Gene: GDB:PRH2  
 A;Cross-references: GDB:119516; OMIM:168790  
 A;Map position: 12p13.2-12p13.2  
 A;Introns: 22/1; 34/1  
 C;Superfamily: proline-rich protein  
 C;Keywords: calcium binding; phosphoprotein; pyroglutamic acid; saliva  
 F;1-16/Domain: signal sequence #status predicted <SIG>  
 F;17-166/Product: protein C #status experimental <PRC>  
 F;17-122/Product: protein A #status experimental <PRA>  
 F;17-45/Region: apatitic mineral binding  
 F;47-71/Product: PRP-3 #status experimental <PRP3>  
 F;123-166/Product: peptide P-C #status experimental  
 F;17/Modified site: pyroglutamic carboxylic acid (Gln) (in mature form) #status experimental  
 F;24,38/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 100.0%; Score 42; DB 1; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 5.7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPPRCR 7  
 Db 118 PRPPRCR 124

RESULT 2  
 B25372

salivary proline-rich phosphoprotein precursor PRH1 (allele PIF) - human  
 C;Species: Homo sapiens (man)  
 C;Date: 29-Aug-1987 #sequence, revision 29-Aug-1987 #text, change 20-Aug-1999  
 C;Accession: B25372; A57868; S02562; G38355; S06153; B27307  
 R;Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
 J. Biol. Chem. 260, 11123-11130, 1985  
 A;Title: Differential RNA splicing and post-translational cleavages in the human salivary proline-rich phosphoprotein precursor PRH1 (allele PIF)  
 A;Reference number: A92492; MUID:85289325; PMID:2993301  
 A;Accession: B25372  
 A;Molecule type: mRNA  
 A;Residues: 1-166 <MAE>  
 A;Cross-references: GB:K03203; NID:9190483; PIDN:AA860184.1; PID:9190484  
 R;Kim, H.S.; Maeda, N.  
 J. Biol. Chem. 261, 6712-6718, 1986  
 A;Title: Structures of two HaeIII-type genes in the human salivary proline-rich phosphoprotein precursor PRH1 (allele PIF)  
 A;Reference number: A57868; MUID:86196106; PMID:3009472  
 A;Accession: A57868  
 A;Molecule type: DNA  
 A;Residues: 1-166 <KIM>  
 A;Cross-references: GB:M13057; NID:9190511; PIDN:AA98807.1; PID:9190512  
 R;Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck  
 Biochem. J. 255, 15-21, 1988  
 A;Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-1 to PRP-6)  
 A;Reference number: S02562; MUID:89061650; PMID:3196309  
 A;Accession: S02562  
 A;Molecule type: protein  
 A;Residues: 47-71 <HAY>  
 R;Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A;Title: Basic proline-rich proteins from human parotid saliva: relationships of the covalent structure to the primary structure  
 A;Reference number: A38355; MUID:91190884; PMID:1849422  
 A;Accession: G38355  
 A;Molecule type: protein  
 A;Residues: 123-166 <KAU>  
 R;Robinson, R.; Kaufman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A;Title: Primary structure and possible origin of the non-glycosylated basic proline-rich protein from human parotid saliva  
 A;Reference number: S06153; MUID:90088384; PMID:2688632  
 A;Accession: S06153  
 A;Molecule type: protein  
 A;Residues: 123-166 <ROB>  
 C;Comment: The proposed biological functions are a highly potent inhibitor of crystal growth  
 C;Genetics:  
 A;Gene: GDB:PRH1  
 A;Cross-references: GDB:119515; OMIM:168730  
 A;Map position: 12p13.2-12p13.2  
 A;Introns: 22/1; 34/1  
 C;Superfamily: proline-rich protein  
 C;Keywords: phosphoprotein; saliva; tandem repeat

Query Match 100.0%; Score 42; DB 2; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 5.7;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPPRCR 7  
 Db 118 PRPPRCR 124

RESULT 3  
 A27307

proline-rich phosphoprotein (gene PRH1, Db allele) - human  
 N;Alternate names: salivary acidic proline-rich protein  
 C;Species: Homo sapiens (man)



C>Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 29-Aug-1997

C/Accession: A27307  
 R/Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A/Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich protein  
 A/Reference number: A27307; MUID:88074309; PMID:3687941  
 A/Accession: A27307  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: DNA  
 A/Residues: 1-171 <AZE>  
 A/Cross-references: EMBL:K03203  
 C/Genetics:  
 A/Gene: GDB:PRH1  
 A/Cross-references: GDB:119515; OMIM:168730  
 A/Map position: 12p13.2-12p13.2  
 C/Superfamily: proline-rich protein  
 C/Keywords: phosphoprotein

Query Match 100.0%; Score 42; DB 2; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 5.9;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
 |||||  
 DB 123 PRPPRGR 129

## RESULT 4

E70610  
 hypothetical protein Rv1215c - Mycobacterium tuberculosis (strain H37RV)  
 C/Species: Mycobacterium tuberculosis  
 C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C/Accession: E70610  
 R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, K.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.  
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A/Reference number: A70500; MUID:98295987; PMID:9634230  
 A/Accession: E70610  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-561 <COL>  
 A/Cross-references: GB:Z93777; GB:AL123456; NID:G3261726; PIDN:CAB07817.1; PID:e311160;  
 A/Experimental source: strain H37RV  
 C/Genetics:  
 A/Gene: Rv1215c

Query Match 100.0%; Score 42; DB 2; Length 561;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
 |||||  
 DB 535 PRPPRGR 541

## RESULT 5

S55609  
 hypothetical protein 13 - equine herpesvirus 2  
 C/Species: equine herpesvirus 2  
 C/Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 08-Oct-1999  
 C/Accession: S55609  
 R/Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.  
 J. Mol. Biol. 249, 520-528, 1995  
 A/Title: The DNA sequence of equine herpesvirus 2.  
 A/Reference number: S55594; MUID:95302501; PMID:7783207  
 A/Accession: S55609  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-200 <TEL>  
 A/Cross-references: GB:U20824; NID:G6951172; PIDN:AAC13802.1; PID:G695187

A/Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

Query Match 92.9%; Score 39; DB 2; Length 200;  
 Best Local Similarity 85.7%; Pred. No. 20;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
 |||||  
 DB 13 PRPPRGR 19

## RESULT 6

AB2338  
 nitrotriacetate monooxygenase, component A Atu6084 [imported] - Agrobacterium tumefaciens  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C/Accession: AB2338  
 R/Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, E.; Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McCle  
 Science 294, 2317-2323, 2001  
 A/Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm  
 ster, E.W.  
 A/Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A/Reference number: AB2577; MUID:21608550; PMID:11743193  
 A/Accession: AB2338  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-450 <KUP>  
 A/Cross-references: GB:AE008690; PIDN:AL46320.1; PID:G17744106; GSPDB:GN00189  
 A/Experimental source: strain C58 (Dupont)  
 C/Genetics:  
 A/Gene: Atu6084  
 C/Superfamily: nitrotriacetate monooxygenase

Query Match 90.5%; Score 38; DB 2; Length 450;  
 Best Local Similarity 85.7%; Pred. No. 62;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
 |||||  
 DB 206 PRPPRGR 212

## RESULT 7

D72579  
 hypothetical protein APE1916 - Aeropyrum pernix (strain K1)  
 C/Species: Aeropyrum pernix  
 C/Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 C/Accession: D72579  
 R/Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Tak  
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;  
 DNA Res. 6, 83-101, 1999  
 A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aerop  
 A/Reference number: A72450; MUID:99310339; PMID:10382966  
 A/Accession: D72579  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-123 <KAW>  
 A/Cross-references: DDBJ:AF000062; NID:G5105244; PIDN:BAA80921.1; PID:dl044707; PID:G5  
 A/Experimental source: strain K1  
 C/Genetics:  
 A/Gene: APE1916

## RESULT 7

D72579  
 hypothetical protein APE1916 - Aeropyrum pernix (strain K1)  
 C/Species: Aeropyrum pernix  
 C/Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 C/Accession: D72579  
 R/Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Tak  
 awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;  
 DNA Res. 6, 83-101, 1999  
 A/Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aerop  
 A/Reference number: A72450; MUID:99310339; PMID:10382966  
 A/Accession: D72579  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-123 <KAW>  
 A/Cross-references: DDBJ:AF000062; NID:G5105244; PIDN:BAA80921.1; PID:dl044707; PID:G5  
 A/Experimental source: strain K1  
 C/Genetics:  
 A/Gene: APE1916

Query Match 88.1%; Score 37; DB 2; Length 123;  
 Best Local Similarity 100.0%; Pred. No. 27;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 6  
 |||||  
 DB 31 PRPPRGR 36

## RESULT 8

A88637  
 A:Title: protein W09G12.9 [imported] - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 10-May-2001  
 C:Accession: A88637  
 R:Anonymous, The C. elegans Sequencing Consortium.  
 Science 282, 2012-2018, 1998  
 A:Title: Genome sequence of the nematode *C. elegans*: a platform for investigating biology  
 A:Reference number: A75000; MUID:9069613; PMID:9851916  
 A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C\_elegans/  
 A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
 A:Accession: A88637  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-137 <STO>  
 A:Cross-references: GB:chr\_IV; PIDN:AA04452.1; PID:G2911887; GSPDB:GN00022; CESP:W09G12  
 C:Genetics:  
 A:Gene: W09G12.9  
 A:Map position: 4

Query Match 88.1%; Score 37; DB 2; Length 137;

Best Local Similarity 100.0%; Pred. No. 30;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPRG 6

DB 120 PRPRG 125

## RESULT 9

H82546  
 A:Title: hypothetical protein XF2514 [imported] - *Xylella fastidiosa* (strain 9a5c)  
 C:Species: *Xylella fastidiosa*  
 C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
 C:Accession: H82546  
 R:Anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing  
 Nature 406, 151-157, 2000  
 A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.  
 A:Reference number: A82515; MUID:20365717; PMID:10910347  
 A:Note: for a complete list of authors see reference number A59328 below  
 A:Accession: H82546  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-151 <SIM>  
 A:Cross-references: GB:AE004059; GB:AE003849; NID:99107718; PIDN:AAF5312.1; GSPDB:GN001  
 A:Experimental source: strain 9a5c  
 R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, P.A.; Acencio, M.; Alvarenga, R.; A  
 Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H  
 as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.  
 submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm  
 J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig  
 Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E  
 A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;  
 F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A  
 Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak  
 A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir  
 M.; Tshukako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z  
 A:Reference number: A59328  
 A:Contents: annotation  
 C:Genetics:  
 A:Gene: XF2514

Query Match 88.1%; Score 37; DB 2; Length 151;

Best Local Similarity 100.0%; Pred. No. 33;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPRG 6

DB 47 PRPRG 52

## RESULT 10

D40595  
 A:Title: hypothetical protein (muta 5' region) - *Streptomyces cinnamonensis*  
 C:Species: *Streptomyces cinnamonensis*  
 C>Date: 03-May-1994 #sequence\_revision 03-May-1994 #text\_change 22-Oct-1999  
 C:Accession: D40595  
 R:Birch, A.; Leiser, A.; Robinson, J.A.  
 J. Bacteriol. 175, 3511-3519, 1993  
 A:Title: Cloning, sequencing, and expression of the gene encoding methylmalonyl-coenzy  
 A:Reference number: A40595; MUID:93273720; PMID:8099072  
 A:Accession: D40595  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-237 <BIR>  
 A:Cross-references: GB:L10064; NID:G153364; PIDN:AAA03039.1; PID:G153365

Query Match 88.1%; Score 37; DB 2; Length 237;

Best Local Similarity 100.0%; Pred. No. 50;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPRG 6

DB 113 PRPRG 118

## RESULT 11

A24026  
 A:Title: erythromycin resistance protein - *Arthrobacter* sp.  
 C:Species: *Arthrobacter* sp.  
 C>Date: 22-Jul-1987 #sequence\_revision 22-Jul-1987 #text\_change 18-Jun-1999  
 C:Accession: A24026  
 R:Roberts, A.N.; Hudson, G.S.; Brenner, S.  
 Gene 35, 259-270, 1985  
 A:Reference number: A24026; MUID:86006275; PMID:4043733  
 A:Accession: A24026  
 A:Molecule type: DNA  
 A:Residues: 1-340 <ROB>  
 A:Cross-references: GB:M11276; NID:G142203; PIDN:AAA22075.1; PID:G142204  
 A:Superfamily: tRNA (adenine-N6-) methyltransferase  
 C:Keywords: antibiotic resistance

Query Match 88.1%; Score 37; DB 2; Length 340;

Best Local Similarity 100.0%; Pred. No. 70;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPRG 6

DB 332 PRPRG 337

## RESULT 12

T32737  
 A:Title: hypothetical protein F54D7.3 - *Caenorhabditis elegans*  
 C:Species: *Caenorhabditis elegans*  
 C>Date: 23-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
 C:Accession: T32737  
 R:Dante, M.; Twyman, B.  
 submitted to the EMBL Data Library, December 1997  
 A:Description: The sequence of *C. elegans* cosmid F54D7.  
 A:Reference number: Z21218  
 A:Accession: T32737  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-401 <DAN>  
 A:Cross-references: EMBL:AF039712; PIDN:AAB96717.1; GSPDB:GN00019; CESP:F54D7.3  
 A:Experimental source: strain Bristol N2; clone F54D7  
 C:Genetics:  
 A:Gene: CESP:F54D7.3  
 A:Map position: 1  
 A:Introns: 47/2; 92/3; 184/3; 228/1; 342/2  
 C:Superfamily: oxytocin receptor

Query Match 88.1%; Score 37; DB 2; Length 401;  
 Best Local Similarity 100.0%; Pred. No. 81;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRG 6  
 |||||  
 Db 370 PRPPRG 375

## RESULT 13

POLJCH  
 Probable gag polyprotein - human T-cell lymphotropic virus type 1  
 C:Species: human T-cell lymphotropic virus type 1, HTLV-1  
 A:Note: host Homo sapiens (man)  
 C:Date: 14-Nov-1983 #sequence\_revision 27-Nov-1985 #text\_change 26-Apr-1996  
 C:Accession: B93954; A91315; A03943  
 R:Seiki, M.; Hattori, S.; Hirayama, Y.; Yoshida, M.  
 Proc. Natl. Acad. Sci. U.S.A. 80, 3618-3622, 1983  
 A:Title: Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus  
 A:Reference number: A93954; MUID:83221647; PMID:6304725  
 A:Accession: B93954  
 A:Molecule type: DNA  
 A:Residues: 1-429 <SEI>  
 A:Experimental source: strain ATK  
 R:Copeland, T.D.; Orszan, S.; Kalyanaraman, V.S.; Sarngadharan, M.G.; Gallo, R.C.  
 FEBS Lett. 162, 390-395, 1983  
 A:Title: Complete amino acid sequence of human T-cell leukemia virus structural protein  
 A:Reference number: A91315; MUID:84029174; PMID:6313426  
 A:Accession: A91315  
 A:Molecule type: protein  
 A:Residues: 345-415, 'T', 417-429 <COP>  
 C:Comment: This protein is synthesized as a gag-pol polyprotein.  
 C:Genetics:

Query Match 88.1%; Score 37; DB 1; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRG 6  
 |||||  
 Db 13 PRPPRG 18

## RESULT 14

POLJCN  
 gag polyprotein - human T-cell lymphotropic virus type 1 (isolate Caribbean)  
 N:Contains: major core protein p19; major core protein p24; nucleic acid-binding protein  
 C:Species: human T-cell lymphotropic virus type 1, HTLV-1  
 A:Note: host Homo sapiens (man)  
 C:Date: 30-Sep-1989 #sequence\_revision 30-Sep-1989 #text\_change 16-Jun-2000  
 C:Accession: A28136  
 R:Malik, K.T.A.; Even, J.; Karpas, A.  
 J. Gen. Virol. 69, 1695-1710, 1988  
 A:Title: Molecular cloning and complete nucleotide sequence of an adult T cell leukaemia  
 bers of the ATL/HTLV-I subgroup.  
 A:Reference number: A92797; MUID:88274338; PMID:2899128  
 A:Accession: A28136  
 A:Molecule type: DNA  
 A:Residues: 1-429 <MAL>  
 A:Cross-references: GB:D13784; GB:D00294; NID:G221866; PIDN:BAA02929.1; PID:G221867  
 C:Genetics:  
 A:Gene: gag  
 C:Superfamily: mammalian retrovirus gag polyprotein II  
 C:Keywords: core protein; polyprotein  
 F:1-130/Product: major core protein p19 #status predicted <P19>  
 F:131-344/Product: major core protein p24 #status predicted <P24>  
 F:345-429/Product: nucleic acid-binding protein p15 #status predicted <P15>

Query Match 88.1%; Score 37; DB 1; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRG 6  
 |||||  
 Db 13 PRPPRG 18

## RESULT 15

S06073  
 gag polyprotein - human T-cell lymphotropic virus type 1  
 C:Species: human T-cell lymphotropic virus type 1, HTLV-1  
 C:Date: 28-Feb-1990 #sequence\_revision 28-Feb-1990 #text\_change 26-Aug-1999  
 C:Accession: S06073  
 R:Gray, G.S.; Bartman, T.; White, M.  
 Nucleic Acids Res. 17, 7998, 1989  
 A:Title: Nucleotide sequence of the core (gag) gene from HTLV-1 isolate MT-2.  
 A:Reference number: S06073; MUID:90016893; PMID:2678008  
 A:Accession: S06073  
 A:Molecule type: mRNA  
 A:Residues: 1-429 <GRA>  
 A:Cross-references: EMBL:X15951; NID:G60425; PIDN:CAA34075.1; PID:G60426  
 C:Superfamily: mammalian retrovirus gag polyprotein II  
 C:Keywords: polyprotein

Query Match 88.1%; Score 37; DB 2; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 86;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRG 6  
 |||||  
 Db 13 PRPPRG 18

Search completed: April 6, 2004, 16:16:54  
 Job time : 8.95047 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:31:19 ; Search time 4.51402 Seconds  
(without alignments)  
80.746 Million cell updates/sec

Title: US-10-009-709-10

Perfect score: 42

Sequence: 1 PRPRGR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	166	PRPC_HUMAN	P02810 homo sapien
2	39	92.9	576	Z384_HUMAN	Q8cf68 homo sapien
3	39	92.9	579	Z384_RAT	Q9eqj4 rattus norv
4	38	90.5	260	DPM1_MOUSE	O70152 mus musculu
5	38	90.5	266	DPM1_CRIGR	Q9w183 cricetulus
6	37	88.1	237	YMU5_STRCM	Q05071 streptomyc
7	37	88.1	340	ERMA_ARTS3	P09891 arthrobacte
8	37	88.1	429	GAG_HTLIA	P03345 human t-cel
9	37	88.1	429	GAG_HTLIC	P14076 human t-cel
10	37	88.1	429	GAG_HTLIM	P14077 human t-cel
11	37	88.1	438	SYH_AERPE	Q9yeb2 aeropyrum p
12	37	88.1	1385	YMS5_CABEL	P34501 caenorhabdi
13	37	88.1	1729	TABP_HUMAN	Q9c0c2 homo sapien
14	36	85.7	447	GASR_HUMAN	P32239 homo sapien
15	36	85.7	450	GASR_FRANA	P30796 praomys nat
16	36	85.7	452	GASR_RABIT	P46627 cryptocolag
17	36	85.7	452	GASR_RAT	P30553 rattus norv
18	36	85.7	453	GASR_CANFA	P30552 canis fami
19	36	85.7	454	GASR_MOUSE	P56481 mus musculu
20	36	85.7	454	GASR_BOVIN	P79266 bos taurus
21	36	85.7	607	MM16_HUMAN	P51512 homo sapien
22	36	85.7	607	MM16_MOUSE	Q9w180 mus musculu
23	36	85.7	607	MM16_RAT	Q35548 rattus norv
24	36	85.7	840	SYFB_STRCO	O88054 streptococ
25	35	83.3	337	TRPD_HALVO	P52562 halobacteri
26	35	83.3	466	VL2_RHPV1	P22165 rhesus papi
27	35	83.3	633	ROR_HUMAN	Q43390 homo sapien
28	35	83.3	641	PXL_STRTO	Q9k194 streptomyc
29	35	83.3	1553	TP2A_CHICK	Q42130 gallus gall
30	34	81.0	178	CDNB_MUSVI	P46529 mustela vis
31	34	81.0	198	CDNB_CRIGR	Q60439 cricetulus
32	34	81.0	198	CDNB_FELCA	O19001 felis silve
33	34	81.0	198	CDNB_HUMAN	P46527 homo sapien

#### ALIGNMENTS

RESULT 1  
PRPC\_HUMAN  
ID PRPC\_HUMAN STANDARD; PRT; 166 AA.  
AC P02810;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-2/PRP-3) (PRP-2/PRP-4) (PIF-F/PIF-S) (Protein A/protein C) [Contains: Peptide P-C].  
DE Peptide P-C.  
GN PRH1 AND PRH2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ALLELES PRH1-4 AND PRH2-1).  
RX MEDLINE=86196106; PubMed=3009472;  
RA Kim H.-S., Maeda N.;  
RT "Structures of two Haell-type genes in the human salivary proline-rich protein multigene family.";  
RL J. Biol. Chem. 261:6712-6718(1986).  
RN [2]  
RP SEQUENCE FROM N.A. (ALLELES PRH2-1 AND PRH1-PIF).  
RX MEDLINE=85289325; PubMed=2993301;  
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
RT "Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system.";  
RL J. Biol. Chem. 260:11123-11130(1985).  
RN [3]  
RP SEQUENCE OF 17-166 (PRP-1 TO PRP-4; PIF-F AND PIF-S).  
RX MEDLINE=89061650; PubMed=3196309;  
RA Hay D.I., Bennick A., Schlesinger D.H., Minaguchi K., Madapallimattam G., Schluckebier S.K.;  
RT "The primary structures of six human salivary acidic proline-rich proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-f).";  
RL Biochem. J. 255:15-21(1988).  
RN [4]  
RP SEQUENCE OF 17-166 FROM N.A. (PA ALLELE).  
RX MEDLINE=88074309; PubMed=3687941;  
RA Azen E.A., Kim H.-S., Goodman E., Flynn S., Maeda N.;  
RT "Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich proteins Pa, Db, and PIF.";  
RL Am. J. Hum. Genet. 41:1035-1047(1987).  
RN [5]  
RP SEQUENCE OF 17-166 (PRP-2).  
RX MEDLINE=86222916; PubMed=3710693;  
RA Schlesinger D.H., Hay D.I.;  
RT "Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of calcium phosphate crystal growth from human parotid saliva.";  
RL Int. J. Pept. Protein Res. 27:373-379(1986).  
RN [6]  
RP SEQUENCE OF 17-166 (PROTEIN C).  
RX MEDLINE=80204368; PubMed=7380845;  
RA Wong R.S.C., Bennick A.;

P55637 rhizobium s  
Q92834 chlamydia p  
Q829f0 escherichia  
Q8f29 escherichia  
P33643 escherichia  
Q917a7 haemophilus  
Q8xgg2 salmonella  
Q8zlv7 versinia pe  
O84436 chlamydia t  
Q9pjw3 chlamydia m  
P17588 herpes simp  
P37998 equus cabal

34 81.0 231 1 Y4RD\_RHISN  
34 81.0 324 1 Y518\_CHLPN  
36 81.0 325 1 RLUD\_ECOL5  
37 81.0 325 1 RLUD\_ECOL6  
38 81.0 325 1 RLUD\_ECOLI  
39 81.0 325 1 RLUD\_HAEDU  
40 81.0 325 1 RLUD\_SALTY  
41 81.0 325 1 RLUD\_VERPE  
42 81.0 329 1 Y425\_CHLTR  
43 81.0 332 1 Y713\_CHLMU  
44 81.0 340 1 LRPI\_HSVIF  
45 81.0 347 1 CD2\_HORSE

"The primary structure of a salivary calcium-binding proline-rich phosphoprotein (protein C), a possible precursor of a related salivary protein A.";  
 J. Biol. Chem. 255:5943-5948 (1980).  
 [7]  
 SEQUENCE OF 17-46 (PROTEIN C).  
 MEDLINE=81191179; PubMed=7228490;  
 RA Schlesinger D.H., Hay D.I.;  
 RA "Primary structure of the active tryptic fragments of human and monkey salivary anionic proline-rich proteins.";  
 Int. J. Pept. Protein Res. 17:34-41 (1981).  
 [8]  
 SEQUENCE OF 17-122 (PROTEIN A).  
 MEDLINE=79173237; PubMed=438215;  
 RA Wong R.S.C., Hofmann T., Bennick A.;  
 RA "The complete primary structure of a proline-rich phosphoprotein from human saliva.";  
 J. Biol. Chem. 254:4800-4808 (1979).  
 [9]  
 SEQUENCE OF 17-122 (PROTEIN A).  
 RA Schlesinger D.H., Hay D.I.;  
 RA "Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent inhibitor of calcium phosphate precipitation in human parotid saliva.";  
 (In) Gross E., Meienhofer J. (eds.);  
 RL peptides: structure and biological function (Proceedings of the 6th American peptide symposium), pp.133-136, Pierce Chemical Co., Rockford IL. (1979).  
 [10]  
 SEQUENCE OF 123-166 (PEPTIDE P-C).  
 MEDLINE=80227634; PubMed=7390979;  
 RA Isemura S., Saichou E., Sanada K.;  
 RA "The amino acid sequence of a salivary proline-rich peptide, P-C, and its relation to a salivary proline-rich phosphoprotein, protein C.";  
 J. Biochem. 87:1071-1077 (1980).  
 [11]  
 RP VARIANT PRH2-3 LYS-163.  
 RA Azen E.A.;  
 RA "A frequent mutation in the acidic proline-rich protein gene, PRH2, causing a Q147K change closely adjacent to the bacterial binding domain of the cognate salivary PRP (Prl) in Afro-Americans.";  
 Hum. Mutat. 12:72-72 (1998).  
 CC -!- FUNCTION: PRP's act as highly potent inhibitors of crystal growth of calcium phosphates. They provide a protective and reparative environment for dental enamel which is important for the integrity of the teeth.  
 CC -!- SUBCELLULAR LOCATION: Secreted.  
 CC -!- PTM: Proteolytically cleaved; PRP-2, PRP-1, and PIF-S yield PRP-4, PRP-3 (protein A), and PIF-F, respectively.  
 CC -!- POLYMORPHISM: Allele PRH1-4 is also known as PRH1 DA allele; allele PRH2-1 is also known as PR1 or protein C; allele PRH2-3 is also known as PR1.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DE EMBL; K03202; AAA60183.1; -  
 DE EMBL; K03203; AAA60184.1; -  
 DE EMBL; M13057; AAA98807.1; -  
 DE EMBL; M13058; AAA98808.1; -  
 DE Genew; HGNC:9366; PRH1.  
 DE Genew; HGNC:9367; PRH2.  
 DE MIM; 168730; -  
 DE MIM; 168790; -  
 DE MIM; 168710; -  
 DE GO; GO:0005615; C:extracellular space; TAS.  
 KW Repeat; Parotid gland; Phosphorylation; Signal; Polymorphism;  
 KW Pyrrrolidone carboxylic acid.

FT SIGNAL 1 16 SALIVARY ACIDIC PROLINE-RICH  
 FT CHAIN 17 166 PHOSPHOPROTEIN 1/3  
 FT CHAIN 17 122 SALIVARY ACIDIC PROLINE-RICH  
 FT CHAIN 123 166 PHOSPHOPROTEIN 3/4.  
 FT DOMAIN 117 46 PEPTIDE P-C.  
 FT MOD RES 17 17 INHIBIT HYDROXYAPATITE FORMATION, BIND  
 FT MOD RES 24 24 TO HYDROXYAPATITE AND CALCIUM.  
 FT MOD RES 38 38 PYRROLIDONE CARBOXYLIC ACID.  
 FT VARIANT 20 20 PHOSPHORYLATION.  
 FT VARIANT 66 66 D->N (in allele PRH1-4).  
 FT VARIANT 163 163 D->N (in allele PRH2-1).  
 FT VARIANT 163 163 Q->K (in allele PRH2-3).  
 FT CONFLICT 41 41 F->P (IN REF. 10).  
 SQ SEQUENCE 166 AA; 17017 MW; A7DF62BF94E3C3EF CRC64;  
 Query Match 100.0%; Score 42; DB 1; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 2.4;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PRPFRGR 7  
 DB 118 PRPFRGR 124  
 RESULT 2  
 Z384 HUMAN STANDARD; PRT; 576 AA.  
 ID Z384 HUMAN  
 AC Q8TF68; O15407; Q8N938;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Zinc finger protein 384 (Nuclear matrix transcription factor 4)  
 DE (CAG repeat protein 1).  
 GN ZNF384 OR NMF4 OR CAGH1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RA Matsuo M.Y.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Brain;  
 RA Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,  
 RA Wakebe H., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,  
 RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,  
 RA Isono Y., Kawai-Hio Y., Saico K., Nishikawa T., Kimura K.,  
 RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
 RA Sugiyama A., Kawakami K., Kanehori K., Fakanashi-Fujii A., Oshima A.,  
 RA Masuho Y., Nagai K., Suzuki Y., Sugano S., Nagahara K.,  
 RA "NEDO human cDNA sequencing project.";  
 RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 395-576 FROM N.A.  
 RC TISSUE=Brain cortex.  
 RX MEDLINE=97369492; PubMed=9225980;  
 RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.-H., Kikwai A.S.,  
 RA Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;  
 RT "cDNAs with long CAG trinucleotide repeats from human brain.";  
 RL Hum. Genet. 100:114-122 (1997).  
 CC -!- FUNCTION: Transcription factor that binds the consensus DNA  
 CC sequence [GC]AAAAA. Seems to bind and regulate the promoters of  
 CC MZF1, MZF3, MZF7 and C/EBP (By similarity).  
 CC -!- SUBUNIT: Interacts with Cas (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).

```

CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q8TF68-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q8TF68-2; Sequence=VSP_006920;
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: Contains 8 C2H2-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB070238; BAB85125.1; -
CC ENBL; AK095734; BAC04618.1; -
CC ENBL; U80736; AAB91437.1; -
CC Genew; HGNC:11955; ZNF384.
CC InterPro; IPR007087; Znf C2H2.
CC Pfam; PF00096; zf-C2H2; 8.
CC ProDom; PD000003; Znf C2H2; 4.
CC SMART; SM00355; Znf C2H2; 8.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; 8.
CC Transcription regulation; zinc-finger; Metal-binding; Nuclear protein;
CC DNA-binding; Repeat; Alternative splicing.
CC ZN_FING 228 250
CC C2H2-TYPE 1.
CC ZN_FING 256 278
CC C2H2-TYPE 2.
CC ZN_FING 284 306
CC C2H2-TYPE 3.
CC ZN_FING 317 339
CC C2H2-TYPE 4.
CC ZN_FING 345 367
CC C2H2-TYPE 5.
CC ZN_FING 373 397
CC C2H2-TYPE 6.
CC ZN_FING 403 425
CC C2H2-TYPE 7.
CC ZN_FING 433 455
CC C2H2-TYPE 8.
CC ZN_FING 461 521
CC GLN-RICH.
CC FT DOMAIN 466 499
CC ALA-RICH.
CC FT VARSPLIC 300 360
CC Missing (in isoform 2).
CC /FTID=VSP_006920.
CC -----
CC SEQUENCE 576 AA; 63091 MW; 2A152786C3C46D90 CRC64;
CC -----
Query Match 92.9%; Score 39; DB 1; Length 576;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC 1 PRPPRGR 7
CC :|||||
CC :|:|:|
CC 188 PRPPRGR 194
CC -----
RESULT 3
ID Z384 RAT STANDARD; PRT; 579 AA.
AC Q9EQJ4; Q9EQJ2; Q9EQJ3; Q9QJ5;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Zinc finger protein 384 (Nuclear matrix transcription factor 4)
DE (Cas-associated zinc finger protein).
DE ZNF384 OR NMP4 OR C12.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1), SUBCELLULAR LOCATION, DNA-BINDING, AND
RP INTERACTION WITH CAS.
RX MEDLINE=20136045; PubMed=10669742;
RA Nakamoto T., Yamagata T., Sakai R., Ogawa S., Honda H., Ueno H.,

```

```

RA Hirano N., Yazaki Y., Hirai H.;
RT "C12, a zinc finger protein that interacts with p130cas and activates
RT the expression of matrix metalloproteinases.";
RL Mol. Cell. Biol. 20:1649-1658(2000).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC STRAIN=Sprague-Dawley;
RX MEDLINE=21024193; PubMed=11149472;
RA Thunyakitpaisal P., Alvarez M., Tokunaga K., Onyia J.E., Hock J.,
RA Choshi N., Feister H., Rhodes S.J., Bidwell J.P.;
RT Cloning and functional analysis of a family of nuclear matrix
RT transcription factors (NP/NMP4) that regulate type I collagen
RT expression in osteoblasts.";
RL J. Bone Miner. Res. 16:10-23(2001).
CC -!- FUNCTION: Transcription factor that binds the consensus DNA
CC sequence [C]AAAAA. Seems to bind and regulate the promoters of
CC MMP1, MMP3, MMP7 and COL1A1.
CC -!- SUBUNIT: Interacts with Cas.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q9EQJ4-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9EQJ4-2; Sequence=VSP_006921;
CC Name=3;
CC IsoId=Q9EQJ4-3; Sequence=VSP_006922;
CC -!- TISSUE SPECIFICITY: Expressed in osteocytes, osteoblasts, and
CC chondrocytes in bone.
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -!- SIMILARITY: Contains 8 C2H2-type zinc fingers.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB019281; BAA89664.1; -
CC ENBL; AF216804; AAG40582.1; -
CC ENBL; AF216805; AAG40583.1; -
CC ENBL; AF216806; AAG40584.1; -
CC HSP; P08153; IZFD.
CC TRANSFAC; T05136; -
CC TRANSFAC; T05137; -
CC TRANSFAC; T05138; -
CC TRANSFAC; T05141; -
CC TRANSFAC; T05142; -
CC InterPro; IPR007087; Znf_C2H2.
CC Pfam; PF00096; zf-C2H2; 8.
CC ProDom; PD000003; Znf C2H2; 4.
CC SMART; SM00355; Znf C2H2; 8.
CC PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
CC PROSITE; PS00157; ZINC_FINGER_C2H2_2; 8.
CC Transcription regulation; zinc-finger; Metal-binding; Nuclear protein;
CC DNA-binding; Repeat; Alternative splicing.
CC ZN_FING 229 251
CC C2H2-TYPE 1.
CC ZN_FING 257 279
CC C2H2-TYPE 2.
CC ZN_FING 285 307
CC C2H2-TYPE 3.
CC ZN_FING 318 340
CC C2H2-TYPE 4.
CC ZN_FING 346 368
CC C2H2-TYPE 5.
CC ZN_FING 374 398
CC C2H2-TYPE 6.
CC ZN_FING 404 426
CC C2H2-TYPE 7.
CC ZN_FING 434 456
CC C2H2-TYPE 8.
CC ZN_FING 462 524
CC GLN-RICH.
CC FT DOMAIN 467 506
CC ALA-RICH.
CC FT VARSPLIC 103 118
CC Missing (in isoform 2).
CC /FTID=VSP_006921.
CC -----
CC VARSPLIC 301 361
CC Missing (in isoform 3).

```

```

FT CONFLICT 178 179 /FTID=VSP 006922.
PT CONFLICT 576 577 CG -> RS (IN REF. 1).
PT CONFLICT 579 AA; 53139 MW; FBC242E0D1050C45 CRC64;
SQ SEQUENCE 579 AA; 53139 MW; FBC242E0D1050C45 CRC64;

Query Match 92.9%; Score 39; DB 1; Length 579;
Best Local Similarity 85.7%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPGR 7
DB 189 PRPPGR 195

RESULT 4
DPM1 MOUSE STANDARD; PRT; 260 AA.
ID AC 070152; Q9D829;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Dolichol-phosphate mannosyltransferase (EC 2.4.1.83) (Dolichol-
phosphate mannosase synthase) (Dolichyl-phosphate beta-D-
mannosyltransferase) (Mannose-P-dolichol synthase) (MPD synthase) (DPM
synthase).
GN DPM1
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98204925; PubMed=9535917;
RA Tomita S., Inoue N., Maeda Y., Ohishi K., Takeda J., Kinoshita T.;
RT "A homologue of Saccharomyces cerevisiae Dm1p is not sufficient for
RT synthesis of dolichol-phosphate-mannose in mammalian cells.";
RL J. Biol. Chem. 273:9249-9254(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Cerebellum, Lung, and Small intestine;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Konno H., Adachi J., Fukuda Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda Y.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Glessi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldairelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: Transfers mannose from GDP-mannose to dolichol
monophosphate to form dolichol phosphate mannose (Dol-P-Man) which
is the mannose donor in pathways leading to N-glycosylation, and O-
glycosyl phosphatidylinositol membrane anchoring, and O-
mannosylation of proteins.
CC -!- CATALYTIC ACTIVITY: GDP-mannose + dolichyl phosphate = GDP +
dolichyl D-mannosyl phosphate.
CC -!- PATHWAY: Glycosylation.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum.
CC -!- SIMILARITY: Belongs to the glycosyltransferase family 2.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

```



KW Transferase; Glycosyltransferase; Endoplasmic reticulum.  
SQ SEQUENCE 266 AA; 29654 MW; 4AFB37EA3AC3329D CRC64;

Query Match 90.5%; Score 38; DB 1; Length 266;  
Best Local Similarity 85.7%; Pred.No. 17;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPPGR 7  
|||||  
Db 20 PRPPQGR 26

RESULT 6  
YMW5\_STRCM STANDARD; PRT; 237 AA.  
ID YMW5\_STRCM  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1994 (Rel. 28, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical 25.0 kDa protein in mutA 5' region (ORF-D).  
OS Streptomyces cinnamonensis  
OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;  
OC Streptomycineae; Streptomycetaceae; Streptomycetes.  
CX NCBI\_TaxID=1500;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3823.5;  
RX MEDLINE=93273720; PubMed=8099072;  
RT "Cloning, sequencing, and expression of the gene encoding  
methylnalonyl-coenzyme A mutase from Streptomyces cinnamonensis.";  
RL J. Bacteriol. 175:3511-3519(1993).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC Query Match 88.1%; Score 37; DB 1; Length 237;  
CC Best Local Similarity 100.0%; Pred.No. 22;  
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPPGR 6  
|||||  
Db 113 PRPPGR 118

RESULT 7  
ERMA\_ARTS3 STANDARD; PRT; 340 AA.  
ID ERMA\_ARTS3  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE rRNA adenine N-6-methyltransferase (EC 2.1.1.48) (Macrolide-  
DE lincosamide-streptogramin B resistance protein) (Erythromycin  
DE resistance protein).  
GN ERMA.  
OS Arthrobacter sp. (strain B3381).  
OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;  
OC Propionibacterineae; Nocardioidaceae; Aeromicrobium.  
CX NCBI\_TaxID=31956;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86006275; PubMed=4043733;

RA Roberts A.N., Hudson G.S., Brenner S.;  
RT "An erythromycin-resistance gene from an erythromycin-producing  
RL strain of Arthrobacter sp.",  
RC Gene 35:259-270(1985).  
CC -!- FUNCTION: THIS PROTEIN PRODUCES A DIMETHYLATION OF THE ADENINE  
CC RESIDUE AT POSITION 2058 IN 23S rRNA, RESULTING IN REDUCED  
CC AFFINITY BETWEEN RIBOSOMES & MACROLIDE-LINCOSAMIDE-STREPTOGRAMIN B  
CC ANTIBIOTICS.  
CC -!- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + rRNA = S-adenosyl-L-  
CC homocysteine + rRNA containing N(6)-methyladenine.  
CC -!- SIMILARITY: Belongs to the rRNA adenine N-6-methyltransferase  
CC family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC ENBL; M1276; AAA22075.1; -  
CC InterPro; IPR001737; RNA\_A\_dimeth.  
CC InterPro; IPR000051; SAM\_Bind.  
CC Pfam; PF00398; RnaAD; 1.  
CC SMART; SM00650; rADC; 1.  
CC PROSITE; PS01131; RNA\_A\_DIMETH; 1.  
CC Antibiotic resistance; Transferase; Methyltransferase.  
CC KW Antitoxic resistance; Transferase; Methyltransferase.  
CC SEQUENCE 340 AA; 37453 MW; E99A714C391952B5 CRC64;  
CC -----  
CC Query Match 88.1%; Score 37; DB 1; Length 340;  
CC Best Local Similarity 100.0%; Pred.No. 32;  
CC Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPPGR 6  
|||||  
Db 332 PRPPGR 337

RESULT 8  
GAG\_HTLIA STANDARD; PRT; 429 AA.  
ID GAG\_HTLIA  
AC P03345;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE GAG polyprotein [Contains: Major core proteins P19 and P24; Nucleic  
DE acid-binding protein P15].  
GN GAG.  
OS Human T-cell leukemia virus type I (strain ATK) (HTLV-I).  
OC Viruses; Retroviridae; Retroviridae; Deltaretrovirus.  
CX NCBI\_TaxID=11926;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=83221647; PubMed=6304725;  
RA Seiki M., Hattori S., Hirayama Y., Yoshida M.;  
RT "Human adult T-cell leukemia virus: complete nucleotide sequence of  
RT the provirus genome integrated in leukemia cell DNA.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:3618-3622(1983).  
RN [2]  
RP SEQUENCE OF 131-155.  
RX MEDLINE=82174582; PubMed=6280175;  
RA Oroszlan S., Sargadharan M.G., Copeland T.D., Kalyanaram V.S.,  
RA Gilden R.V., Gallo R.C.;  
RT "Primary structure analysis of the major internal protein p24 of  
RT human type C T-cell leukemia virus.";  
RL Proc. Natl. Acad. Sci. U.S.A. 79:1291-1294(1982).  
RN [3]  
RP SEQUENCE OF 345-429.  
RX MEDLINE=84029174; PubMed=6313426;  
RA Copeland T.D., Oroszlan S., Kalyanaram V.S., Sargadharan M.G.,  
RA Gallo R.C.;  
RT "Complete amino acid sequence of human T-cell leukemia virus

RT structural protein p15".  
 RL FEBS Lett. 162:390-395(1993).  
 CC -!- PFM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -!- MISCELLANEOUS: This protein is synthesized as a Gag-Pol  
 CC POLYPROTEIN.  
 CC -!- SIMILARITY: Contains 2 CCHC-type zinc fingers.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: J02029; AAA96672.1; -.  
 DR PIR: B93954; FOLJGH.  
 DR PDB: 1QRJ; 26-SEP-01.  
 DR InterPro: IPR003139; Gag\_p19.  
 DR InterPro: IPR000721; Gag\_p24.  
 DR InterPro: IPR008916; Retrov\_Capsid\_C.  
 DR InterPro: IPR001878; Znf\_CCHC.  
 DR Pfam: PF02228; Gag\_p19; 1.  
 DR Pfam: PF00607; Gag\_p24; 1.  
 DR PPRINTS: PR00939; C2HCZNFINGER.  
 DR SMART: SM00343; Znf\_CCHC; 2.  
 DR PROSITE: PS00158; Znf\_CCHC; 1.  
 KW Core protein; Polyprotein; Zinc-finger; Repeat; 3D-structure.  
 FT CHAIN 1 130 MAJOR CORE PROTEIN P19.  
 FT CHAIN 131 344 MAJOR CORE PROTEIN P24.  
 FT CHAIN 345 429 NUCLEIC ACID-BINDING PROTEIN P15.  
 FT ZN\_FING 355 372 CCHC-TYPE 1.  
 FT ZN\_FING 378 395 CCHC-TYPE 2.  
 FT CONFLICT 416 416 A -> T (IN REF. 2); AAA96672).  
 FT CONFLICT 424 424 S -> F (IN REF. 1; AAA96672).  
 SQ SEQUENCE 429 AA; 47496 MW; 03FFD4E5A4500284 CRC64;  
 Query Match 88.1%; Score 37; DB 1; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PRPPRG 6  
 DB 13 PRPPRG 18  
 RESULT 9  
 GAG HTL1C  
 ID GAG HTL1C STANDARD; PRT; 429 AA.  
 AC P14076;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE GAG polyprotein [Contains: Major core proteins P19 and P24; Nucleic  
 DE acid-binding protein p15].  
 GN GAG.  
 OS Human T-cell leukemia virus type I (Caribbean isolate) (HTLV-I).  
 OC Viruses; Retrov. viruses; Retroviridae; Deltaretrovirus.  
 OX NCBI\_TaxID=11927;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88274338; PubMed=2899128;  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE GAG polyprotein [Contains: Major core proteins P19 and P24; Nucleic  
 DE acid-binding protein p15].  
 GN GAG.  
 OS Human T-cell leukemia virus type I (Caribbean isolate) (HTLV-I).  
 OC Viruses; Retrov. viruses; Retroviridae; Deltaretrovirus.  
 OX NCBI\_TaxID=11927;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88274338; PubMed=2899128;  
 RA Malik K.T.A., Even J., Karpas A.;  
 RT "Molecular cloning and complete nucleotide sequence of an adult T  
 RT cell leukaemia virus/human T cell leukaemia virus type I  
 RT (ATLV/HTLV-I) isolate of Caribbean origin: relationship to other  
 RT members of the ATLV/HTLV-I subgroup".  
 RL J. Gen. Virol. 69:1695-1710(1988).  
 CC -!- PFM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -!- SIMILARITY: Contains 2 CCHC-type zinc fingers.  
 CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: D13784; BAA02929.1; -.  
 DR PIR: A28136; FOLJCN.  
 DR InterPro: IPR003139; Gag\_p19.  
 DR InterPro: IPR000721; Gag\_p24.  
 DR InterPro: IPR008916; Retrov\_Capsid\_C.  
 DR InterPro: IPR001878; Znf\_CCHC.  
 DR Pfam: PF02228; Gag\_p19; 1.  
 DR Pfam: PF00607; Gag\_p24; 1.  
 DR PPRINTS: PR00939; C2HCZNFINGER.  
 DR SMART: SM00343; Znf\_CCHC; 2.  
 DR PROSITE: PS00158; Znf\_CCHC; 1.  
 KW Core protein; Polyprotein; Zinc-finger; Repeat.  
 FT CHAIN 1 130 MAJOR CORE PROTEIN P19.  
 FT CHAIN 131 344 MAJOR CORE PROTEIN P24.  
 FT CHAIN 345 429 NUCLEIC ACID-BINDING PROTEIN P15.  
 FT ZN\_FING 355 372 CCHC-TYPE 1.  
 FT ZN\_FING 378 395 CCHC-TYPE 2.  
 SQ SEQUENCE 429 AA; 47514 MW; CFBF686497B738EC CRC64;  
 Query Match 88.1%; Score 37; DB 1; Length 429;  
 Best Local Similarity 100.0%; Pred. No. 41;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 PRPPRG 6  
 DB 13 PRPPRG 18  
 RESULT 10  
 GAG HTL1M  
 ID GAG HTL1M STANDARD; PRT; 429 AA.  
 AC P14077;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE GAG polyprotein [Contains: Major core proteins P19 and P24; Nucleic  
 DE acid-binding protein p15].  
 GN GAG.  
 OS Human T-cell leukemia virus type I (isolate MT-2) (HTLV-I).  
 OC Viruses; Retrov. viruses; Retroviridae; Deltaretrovirus.  
 OX NCBI\_TaxID=11928;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90016893; PubMed=2678008;  
 RA Gray G.S., Bartman T., White M.;  
 RT "Nucleotide sequence of the core (gag) gene from HTLV-1 isolate  
 RT MT-2".  
 RL Nucleic Acids Res. 17:7998-7998(1989).  
 CC -!- PFM: Specific enzymatic cleavages in vivo yield mature proteins.  
 CC -!- SIMILARITY: Contains 2 CCHC-type zinc fingers.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: X15951; CAA34075.1; -.  
 DR PIR: S06073; S06073.  
 DR PDB: 1G03; 21-MAR-01.  
 DR InterPro: IPR003139; Gag\_p19.

```
DR InterPro; IPR000721; Gag_p24.
DR InterPro; IPR008916; Retrov_Capsid_C.
DR InterPro; IPR008919; Retrov_Capsid_N.
DR InterPro; IPR001878; Znf_CCHC.
DR Pfam; PF02228; Gag_p19; 1.
DR Pfam; PF00607; Gag_p24; 1.
DR Pfam; PF00098; Zf_CCHC; 2.
DR PRINTS; PR00939; C2HCZNFINGER.
DR SMART; SM00343; Znf_C2HC; 2.
DR PROSITE; PS0158; ZF_CCHC; 1.
KW Core protein; Polyprotein; Zinc-finger; Repeat; 3D-structure.
FT CHAIN 1 130 MAJOR CORE PROTEIN P19.
FT CHAIN 131 344 MAJOR CORE PROTEIN P24.
FT CHAIN 345 429 NUCLEIC ACID-BINDING PROTEIN P15.
FT CHAIN 355 372 CCHC-TYPE 1.
FT ZN_FING 378 395 CCHC-TYPE 2.
FT ZN_FING 378 395 CCHC-TYPE 2.
SQ SEQUENCE 429 AA; 47584 MW; EF5201C934EF0291 CRC64;

Query Match 88.1%; Score 37; DB 1; Length 429;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPPRG 6
Db 13 PRPPRG 18

RESULT 11
ID SYH_AERPE STANDARD; PRT; 438 AA.
AC Q9YEB2;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Histidyl-tRNA synthetase (EC 6.1.1.21) (Histidine-tRNA ligase)
DE (HISRS).
GN HISRS OR APE0662.
OS Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococcales; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=993101339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyana A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kusida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
CC -!- CATALYTIC ACTIVITY: ATP + L-histidine + tRNA(His) = AMP +
CC diphosphate + L-histidyl-tRNA(His).
CC -!- SUBCELLULAR LOCATION: Cytoplasm;c.
CC -!- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF000060; BAA79634.1; -.
CC PIR; B72654; B72654.
CC HSSP; O32422; IQ50.
CC HAMAP; MF_00127; -. 1.
CC InterPro; IPR004154; HGTP_anticondon.
CC InterPro; IPR004516; HisS.
```

```
DR InterPro; IPR002314; tRNA-synt 2b.
DR InterPro; IPR006155; tRNA_ligase-II.
DR Pfam; PF03129; HGTP_anticondon; 1.
DR Pfam; PF00587; tRNA-synt 2b; 1.
DR TIGRFAMS; TIGR00442; hisS; 1.
DR PROSITE; PS00862; AA tRNA_LIGASE II; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
SQ SEQUENCE 438 AA; 49111 MW; ECCE67F59A9FB7BC CRC64;

Query Match 88.1%; Score 37; DB 1; Length 438;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPPRG 6
Db 9 PRPPRG 14

RESULT 12
ID YMS5_CABEL STANDARD; PRT; 1385 AA.
AC P34501;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein K03H1.5 in chromosome III.
GN K03H1.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=94119718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M., Coulson A.,
RA Bonfield J., Burton J., Connell M., Copsey I., Cooper J., Fraser A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laiister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkneen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaubin M., Vaughan K.,
RA Waterston R., Watson A., Weinstein L., Wilkinson-Sproat J.,
RA Wooldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38(1994).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z29560; CAA82664.1; -.
CC PIR; H88569; H88569.
CC WormPep; K03H1.5; CE03459.
CC InterPro; IPR005533; AMOP.
CC InterPro; IPR002909; IPT_TIG.
CC InterPro; IPR003886; Nidogen_ext.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC InterPro; IPR001846; VWF_d.
CC Pfam; PF03782; AMOP; 1.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF00094; vwd; 1.
CC SMART; SM00723; AMOP; 1.
CC SMART; SM00032; CCP; 1.
CC SMART; SM00429; IPT; 1.
```

```
DR SMART; SM00539; NIDO; 1.
DR SMART; SM00216; VWD; 1.
DR Hypothetical protein.
SQ SEQUENCE 1385 AA; 159181 MW; BDCD8F59CEA38C03 CRC64;
Query Match
88.1%; Score 37; DB 1; Length 1385;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PRPPRG 6
Db 1371 PRPPRG 1376
RESULT 13
TASP_HUMAN
ID TASP_HUMAN STANDARD; PRT; 1729 AA.
AC Q9COC2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 182 kDa tankyrase 1-binding protein.
GN TNKS1BP1 OR TAB182 OR KIAA1741.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta, and Testis;
RX MEDLINE=21950796; PubMed=11854288;
RA Seimiya H., Smith S.;
RT "The telomeric poly(ADP-ribose) polymerase, tankyrase 1, contains
RT multiple binding sites for telomeric repeat binding factor 1 (TRF1)
RT and a novel acceptor. 182-kDa tankyrase-binding protein (TAB182).";
RL J. Biol. Chem. 277:14116-14126(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21082932; PubMed=11214970;
RA Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIX.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 7:347-355(2000).
RN [3]
RP SEQUENCE OF 495-1729 FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=22579292; PubMed=12693554;
RA Jikuya H., Takano J., Kikuno R., Hirosewa M., Nagase T., Nomura N.,
RA Ohara O.;
RT "Characterization of long cDNA clones from human adult spleen. II. The
RT complete sequences of 81 cDNA clones.";
RL DNA Res. 10:49-57(2003).
CC -1- SUBUNIT: Binds to the ANK repeat domain of TNKS1 and TNKS2.
CC -1- SUBCELLULAR LOCATION: Nuclear and cytoplasmic. Colocalizes with
CC chromosomes during mitosis, and in the cytoplasm with cortical
CC actin.
CC -1- TISSUE SPECIFICITY: Detected in testis, ovary, lung, skeletal
CC muscle, heart, prostate and pancreas, and at very low levels in
CC brain and peripheral blood leukocytes.
CC -1- PTM: ADP-ribosylated by TNKS1 (in vitro).
CC -1- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in position 1071, 1097 and 1467.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF441771; AAM15531.1; -.
DR EMBL; AB051528; BAB21832.2; ALT INIT.
DR EMBL; AK071413; BAB84939.1; ALT_FRAME.
DR Genew; HGNC:19081; TNKS1BP1.
DR GK; Q9COC2; -.
DR MIM; 607104; -.
DR GO; GO:0005737; C:cytoplasm; NAS.
DR GO; GO:0005724; C:nuclear telomeric heterochromatin; NAS.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0030506; F:ankyrin binding; NAS.
DR GO; GO:0019899; F:enzyme binding; NAS.
DR GO; GO:0007004; P:telomerase-dependent telomere maintenance; NAS.
DR InterPro; IPR008979; Gal bind like.
KW Nuclear protein; Chromosomal protein; ADP-ribosylation.
FT DOMAIN 2 103 ARG/GLU/LYS/PRO-RICH (CHARGED).
FT DOMAIN 127 767 PRO-RICH.
FT DOMAIN 210 1572 ACIDIC.
FT DOMAIN 1010 1340 GLY-RICH.
FT DOMAIN 1450 1542 TANKYRASE-BINDING.
FT DOMAIN 1572 1729 ARG/GLU/LYS-RICH (CHARGED).
FT DOMAIN 1629 1635 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 1723 1729 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT CONFLICT 84 84 P -> L (IN REF. 2).
FT CONFLICT 322 322 S -> T (IN REF. 2).
FT CONFLICT 388 388 S -> P (IN REF. 2).
FT CONFLICT 554 554 Q -> H (IN REF. 3).
FT CONFLICT 604 604 P -> S (IN REF. 1).
FT CONFLICT 1450 1450 F -> S (IN REF. 2).
SQ SEQUENCE 1729 AA; 181814 MW; C65F38FA37045C4A CRC64;
Query Match
88.1%; Score 37; DB 1; Length 1729;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PRPPRG 6
Db 65 PRPPRG 70
RESULT 14
GASR_HUMAN
ID GASR_HUMAN STANDARD; PRT; 447 AA.
AC P32239;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Gastrin/cholecystokinin type B receptor (CCK-B receptor) (CCK-BR).
GN CCKBR OR CCKBR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]_
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93216795; PubMed=7681836;
RA Lee Y.-M., Beinborn M., McBride E.W., Lu M., Kolakowski L.F. Jr.,
RA Kopin A.S.;
RT "The human brain cholecystokinin-B/gastrin receptor. Cloning and
RT characterization.";
RL J. Biol. Chem. 268:8164-8169(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93080572; PubMed=1280419;
RA Pilegna J.R., de Weerth A., Huppi K., Wank S.A.;
RT "Molecular cloning of the human brain and gastric cholecystokinin
RT receptor: structure, functional expression and chromosomal
RT localization.";
RL Biochem. Biophys. Res. Commun. 189:296-303(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
```



CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 CC EMBL; D12817; BAA02250.1; -  
 DR PIR; JQ1614;  
 DR InterPro; IPR000276; GPCR\_Rhodopsin.  
 DR Pfam; PF00001; 7tm.1; 1.  
 DR PRINTS; PR00337; GPCRHOPOPSN  
 DR PROSITE; PS00237; G-PROTEIN\_RECEPT\_F1\_1; 1.  
 DR PROSITE; PS0262; G-PROTEIN\_RECEPT\_F1\_2; 1.  
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;  
 KW Lipoprotein; Palmitate.  
 FT DOMAIN 1 57 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 58 79 1 (POTENTIAL).  
 FT DOMAIN 80 87 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 88 109 2 (POTENTIAL).  
 FT DOMAIN 110 131 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 132 150 3 (POTENTIAL).  
 FT DOMAIN 151 170 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 171 189 4 (POTENTIAL).  
 FT DOMAIN 190 219 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 220 242 5 (POTENTIAL).  
 FT DOMAIN 243 336 CYTOPLASMIC (POTENTIAL).  
 FT TRANSMEM 337 358 6 (POTENTIAL).  
 FT DOMAIN 359 376 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 377 397 7 (POTENTIAL).  
 FT DOMAIN 398 450 CYTOPLASMIC (POTENTIAL).  
 FT CARBOHYD 7 7 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT DISULFID 36 36 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT LIPID 127 205 BY SIMILARITY.  
 FT LIPID 411 411 S-palmitoyl cysteine (By similarity).  
 SQ SEQUENCE 450 AA; 48755 MW; EF13B8B5FAA857A CRC64;

Query Match 85.7%; Score 36; DB 1; Length 450;  
 Best Local Similarity 85.7%; Pred. No. 63;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PRPRGR 7  
 |||||  
 Db 413 PRPRAR 419

Search completed: April 6, 2004, 16:08:00  
 Job time : 4.51402 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:51:34 ; Search time 25.3178 Seconds  
(without alignments)  
87.236 Million cell updates/sec

Title: US-10-009-709-10

Perfect score: 42

Sequence: 1 PRPPRGR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- SPTREMBL 25:
- 1: sp\_archaea:
  - 2: sp\_bacteria:
  - 3: sp\_fungi:
  - 4: sp\_human:
  - 5: sp\_invertebrate:
  - 6: sp\_mammal:
  - 7: sp\_mbc:
  - 8: sp\_organella:
  - 9: sp\_phage:
  - 10: sp\_plant:
  - 11: sp\_rodent:
  - 12: sp\_virus:
  - 13: sp\_vertebrate:
  - 14: sp\_unclassified:
  - 15: sp\_virus:
  - 16: sp\_bacteriophage:
  - 17: sp\_archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	332	Q54729	Q54729 synchococc
2	42	100.0	535	Q84SD0	Q84SD0 oryza sativ
3	42	100.0	561	Q05316	Q05316 mycobacteri
4	42	100.0	561	Q7U0D5	Q7U0D5 mycobacteri
5	39	92.9	200	Q66619	Q66619 equine herp
6	39	92.9	397	Q9EQJ1	Q9EQJ1 rattus norv
7	39	92.9	460	Q7Z722	Q7Z722 homo sapien
8	39	92.9	517	Q8C3E0	Q8C3E0 mus musculu
9	38	90.5	450	Q8U651	Q8U651 agrobacteri
10	38	90.5	2406	Q9BZS0	Q9BZS0 homo sapien
11	38	90.5	2414	Q9HCL7	Q9HCL7 homo sapien
12	37	88.1	72	Q86918	Q86918 human t-lym
13	37	88.1	117	Q9K263	Q9K263 streptomyce
14	37	88.1	118	Q9XS00	Q9XS00 canis faml
15	37	88.1	123	Q9YAW7	Q9YAW7 aeropyrum p
16	37	88.1	129	Q9IZL0	Q9IZL0 simian t-ly

17	37	88.1	136	4	Q9NWG6	Q9NWG6 homo sapien
18	37	88.1	137	5	O45200	O45200 caenorhabdi
19	37	88.1	151	16	Q9FAK1	Q9FAK1 xyella fas
20	37	88.1	202	15	Q82444	Q82444 human t-lym
21	37	88.1	219	10	Q9LSN5	Q9LSN5 arabidopsis
22	37	88.1	247	4	Q8NIW5	Q8NIW5 homo sapien
23	37	88.1	322	11	Q7TNE5	Q7TNE5 mus musculu
24	37	88.1	401	5	O44731	O44731 caenorhabdi
25	37	88.1	407	10	Q8H3U2	Q8H3U2 oryza sativ
26	37	88.1	428	15	Q9WS55	Q9WS55 simian t-ly
27	37	88.1	429	15	Q82230	Q82230 human t-lym
28	37	88.1	429	15	O56621	O56621 human t-lym
29	37	88.1	429	15	O56229	O56229 human t-lym
30	37	88.1	429	15	Q90041	Q90041 human t-lym
31	37	88.1	449	10	Q7X8A2	Q7X8A2 oryza sativ
32	37	88.1	449	15	Q9QRA0	Q9QRA0 human t-lym
33	37	88.1	474	11	Q61120	Q61120 mus musculu
34	37	88.1	513	16	Q9A731	Q9A731 caulobacter
35	37	88.1	590	12	Q7TFS2	Q7TFS2 thesus cyto
36	37	88.1	594	4	Q8TAP2	Q8TAP2 homo sapien
37	37	88.1	594	4	Q92529	Q92529 homo sapien
38	37	88.1	594	11	O70143	O70143 rattus norv
39	37	88.1	647	15	Q9Q899	Q9Q899 human t-lym
40	37	88.1	651	15	O58227	O58227 human t-lym
41	37	88.1	679	4	Q96RU3	Q96RU3 homo sapien
42	37	88.1	686	10	Q8RZZ7	Q8RZZ7 oryza sativ
43	37	88.1	839	4	Q96IE7	Q96IE7 homo sapien
44	37	88.1	1179	4	Q13545	Q13545 homo sapien
45	37	88.1	1273	15	Q9QRA1	Q9QRA1 human t-lym

## ALIGNMENTS

### RESULT 1

Q54729 PRELIMINARY; PRT; 332 AA.  
AC Q54729;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
DE ORF11692 (Fragment).  
OS Synchococcus sp. (strain PCC 7942) (Anacystis nidulans R2).  
OC Bacteria; Cyanobacteria; Chroococcales; Synchococcus.  
OX NCBI\_taxid=1140;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PCC 7942;  
RA Tsinoremas N.F., Golden S.S.;  
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
DR EMBL; U4761; AAA8664.1; -;  
DR InterPro; IPR001064; CRYSTALLIN.  
DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
FT NON TER 1  
SQ SEQUENCE 332 AA; 38574 MW; CA17B5F4B0F2ED09 CRC64;

Query Match 100.0%; Score 42; DB 2; Length 332;  
Best Local Similarity 100.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7

Db 282 PRPPRGR 288

### RESULT 2

Q84SD0 PRELIMINARY; PRT; 535 AA.  
ID Q84SD0;  
AC Q84SD0;  
DT 01-JUN-2003 (TREMBlrel. 24, Created)  
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE P0577B11.16 protein.

GN P0577B11.16.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Katayose Y.;  
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, PAC  
 clone-P0577B11".  
 RL Submitted (JUL-2002) to the ENBL/GenBank/DBJ databases.  
 DR EMBL; AF005504; BAC57819.1; -;  
 DR InterPro; IPR002885; PPR; 8.  
 DR Pfam; PF01535; PPR; 8.  
 DR TIGRFAMs; TIGR00756; PPR; 9.  
 SQ SEQUENCE 535 AA; 58557 MW; C2712C4178582982 CRC64;

Query Match 100.0%; Score 42; DB 10; Length 535;  
 Best Local Similarity 100.0%; Pred. No. 23; Indels 0; Gaps 0;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PRPPRGR 7  
 |||||  
 DB 114 PRPPRGR 120

RESULT 3  
 O05316 PRELIMINARY; PRT; 561 AA.  
 ID O05316;  
 AC O05316;  
 DT 01-JUL-1997 (TRENBLrel. 04, Created)  
 DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Hypothetical protein (Diester hydrolase, putative).  
 GN RV1215C OR MTC1364.27C OR MT1253.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37Rv;  
 RX MEDLINE=98295987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
 Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;  
 RT "Deciphering the biology of Mycobacterium tuberculosis from the  
 complete genome sequence."  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 Delcher A., Utterback T., Weidman J., Knouri H., Gill J., Mikula A.,  
 Bishai W.;  
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 laboratory strains."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z93777; CAB07817.1; -;  
 DR EMBL; AE007002; AX45510.1; -;  
 DR PIR; E70610; E70610.  
 DR TIGR; MT1253; -;  
 DR TuberculList; RV1215C; -;  
 GO; GO:0005634; C:nucleus; IEA.

GO; GO:0003677; F:DNA binding; IEA.  
 GO; GO:0016787; F:hydrolase activity; IEA.  
 GO; GO:0008451; F:Kaa-Pro aminopeptidase activity; IEA.  
 GO; GO:0008152; P:metabolism; IEA.  
 GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR005674; CoeB\_NonD.  
 DR InterPro; IPR008979; Gal\_Bind\_Like.  
 DR InterPro; IPR001005; Myb\_DNA\_Binding.  
 DR InterPro; IPR000383; Peptidase\_S15.  
 DR Pfam; PF02129; Peptidase\_S15; 1.  
 DR TIGRFAMs; TIGR00976; /NonD; 1.  
 DR PROSITE; PS00037; MYB\_1; 1.  
 KW Hypothetical protein; Hydrolase; Complete proteome.  
 SQ SEQUENCE 561 AA; 62610 MW; 6D2C3253F2D3598D CRC64;

Query Match 100.0%; Score 42; DB 16; Length 561;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PRPPRGR 7  
 |||||  
 DB 535 PRPPRGR 541

RESULT 4  
 Q7U0D5 PRELIMINARY; PRT; 561 AA.  
 ID Q7U0D5;  
 AC Q7U0D5;  
 DT 01-OCT-2003 (TRENBLrel. 25, Created)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)  
 DE Conserved hypothetical protein.  
 GN MB1247C.  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AF2122/97;  
 RX MEDLINE=22709107; PubMed=12788972;  
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,  
 Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,  
 Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
 RT "The complete genome sequence of Mycobacterium bovis."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).  
 KW Complete proteome.  
 SQ SEQUENCE 561 AA; 62626 MW; C35176E8172866AD CRC64;

Query Match 100.0%; Score 42; DB 16; Length 561;  
 Best Local Similarity 100.0%; Pred. No. 24;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PRPPRGR 7  
 |||||  
 DB 535 PRPPRGR 541

RESULT 5  
 Q66619 PRELIMINARY; PRT; 200 AA.  
 ID Q66619;  
 AC Q66619;  
 DT 01-NOV-1996 (TRENBLrel. 01, Created)  
 DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
 DE ORF 13.  
 OS Equine herpesvirus type 2 (strain 86/87) (EHV-2).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Gammaherpesvirinae.  
 OX NCBI\_TaxID=82831;  
 RN [1]



```

RP SEQUENCE FROM N.A.
RX MEDLINE=95302501; PubMed=7783207;
RA Telford E.A.R.;
RT "The DNA sequence of equine herpesvirus 2.";
RN J. Mol. Biol. 249:520-528(1995).
[2]
RP SEQUENCE FROM N.A.
RA Telford E.A.R.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U20824; AAC13802.1; -.
DR PIR; S55609; S55609.
SQ SEQUENCE 200 AA; 22356 MW; CF3306CF3C0974B1 CRC64;

Query Match          92.9%; Score 39; DB 12; Length 200;
Best Local Similarity 85.7%; Pred. No. 30;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPGR 7
Db 13 PKPGR 19

RESULT 6
Q9EQJ1 PRELIMINARY; PRT; 397 AA.
ID Q9EQJ1
AC Q9EQJ1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Nuclear matrix transcription factor.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX STEIN-Sprague-Pawley;
RA MEDLINE=102419; PubMed=11149472;
RA Thunyakitpisal P., Alvarez M., Tokunaga K., Onyia J.E., Hock J.,
RA Ohashi N., Feister H., Rhodes S.J., Bidwell J.P.;
RT "Cloning and functional analysis of a family of nuclear matrix
RT transcription factors (NP/NM4) that regulate type I collagen
RT expression in osteoblasts.";
RL J Bone Miner. Res. 16:10-23(2001).
DR EMBL; AF16807; AAC40585.1; -.
DR InterPro; IPR007087; Znf_C2H2.
DR Pfam; PF00096; Zf-C2H2_4.
DR ProDom; PD000003; Znf_C2H2_3.
DR SMART; SMO0355; Znf_C2H2_5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
KW Metal-binding; Zinc; Zinc-finger.
SQ SEQUENCE 397 AA; 43030 MW; C79D0B36AC23E480 CRC64;

Query Match          92.9%; Score 39; DB 11; Length 397;
Best Local Similarity 85.7%; Pred. No. 57;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPGR 7
Db 189 PKPGR 195

RESULT 7
Q7Z722 PRELIMINARY; PRT; 460 AA.
ID Q7Z722
AC Q7Z722;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Testis;
RA MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner J., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Scapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,
RA Bha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McSwain P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettaman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE=Testis;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053361; AAH53361.1; -.
KW Hypothetical protein.
SQ SEQUENCE 460 AA; 50409 MW; 0C01860AFC3C71F2 CRC64;

Query Match          92.9%; Score 39; DB 4; Length 460;
Best Local Similarity 85.7%; Pred. No. 65;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPGR 7
Db 133 PKPGR 139

RESULT 8
Q8C3E0 PRELIMINARY; PRT; 517 AA.
ID Q8C3E0
AC Q8C3E0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Unclassifiable.
GN CL30073D16RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Head;
RA MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK086185; BAC39626.1; -.
DR PIR; PT0649; PT0712.
DR MGD; MGI:2443203; Cl30073D16RIK.
DR InterPro; IPR007087; Znf_C2H2.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 1.
SQ SEQUENCE 517 AA; 55137 MW; 4378A6D9895D9ABC CRC64;

```

Query Match 92.9%; Score 39; DB 11; Length 517;  
 Best Local Similarity 85.7%; Pred. No. 72;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
 Db 189 PKPPRGR 195

## RESULT 9

Q8U651 PRELIMINARY; PRT; 450 AA.  
 ID Q8U651  
 AC Q8U651;  
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Nitroliotriacetate monooxygenase, component A.  
 GN ATU6084 OR AGR\_PTI\_161.  
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).  
 OG Plasmid pTiC58.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.  
 OX NCBI\_TaxID=176299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608550; PubMed=11743193;  
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,  
 RA Chou V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,  
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,  
 RA Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,  
 RA Kuyavlin T., Levy R., Li M.-J., McClelland E., Palmieri A.,  
 RA Raymond C., Rouse G., Saephammachak C., Wu Z., Romero P., Gordon D.,  
 RA Zhang S., Yoo H., Ito Y., Biddle P., Jung M., Krespan W., Perry M.,  
 RA Gordon-Kam B., Liac L., Kim S., Hendrick C., Zhao Z.-X., Dolan M.,  
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,  
 RA Nester E.W.;  
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens  
 C58.";  
 RL Science 294:2317-2323(2001).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21608551; PubMed=11743194;  
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,  
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,  
 RA Roumel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,  
 RA Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,  
 RA Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,  
 RA Cielo C., Slater S.;  
 RT "Genome sequence of the plant pathogen and biotechnology agent  
 Agrobacterium tumefaciens C58.";  
 RL Science 294:2323-2328(2001).  
 DR EMBL; AF009427; AAL46320.1; -;  
 DR EMBL; AF007935; AAK91044.1; -;  
 DR FIR; AB2238; AB2328.  
 DR GO; GO:0046821; C:exochromosomal DNA; IEA.  
 DR GO; GO:0005840; C:ribosome; IEA.  
 DR GO; GO:0004497; F:monooxygenase activity; IEA.  
 DR GO; GO:0003735; F:structural constituent of ribosome; IEA.  
 DR GO; GO:0006412; P:protein biosynthesis; IEA.  
 DR InterPro; IPR002103; B:ac. luciferase.  
 DR InterPro; IPR001865; Ribosomal S2.  
 DR PROSITE; PS00962; RIBOSOMAL\_S2\_1;  
 KW Monooxygenase; Plasmid; Complete proteome.  
 SQ SEQUENCE 450 AA; 50163 MW; D3ADB8261D68C026 CRC64;

Query Match 90.5%; Score 38; DB 16; Length 450;  
 Best Local Similarity 85.7%; Pred. No. 94;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
 Db 206 PRPPRGR 212

## RESULT 10

Q9BZS0 PRELIMINARY; PRT; 2406 AA.  
 ID Q9BZS0  
 AC Q9BZS0;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Kappa B and V(D)J recombination signal sequences binding protein.  
 GN KRC.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21100880; PubMed=11161801;  
 RA Hicar M.D., Liu Y., Allen C.E., Wu L.C.;  
 RA "Structure of the Human Zinc Finger Protein HIVEP3: Molecular Cloning,  
 RT Expression, Exon-Intron Structure, and Comparison with Paralogous  
 RT Genes HIVEP1 and HIVEP2.";  
 RL Genomics 71:189-190(2001).  
 DR EMBL; AF278765; AAK01082.1; -;  
 DR HSP; P15822; 1BBO.  
 DR GO; GO:0005634; C:nucleus; NAS.  
 DR GO; GO:0003700; F:transcription factor activity; TAS.  
 DR GO; GO:0006355; F:regulation of transcription, DNA-dependent; TAS.  
 DR InterPro; IPR02034; AIPM/Hcit\_synth.  
 DR InterPro; IPR007087; Znf\_C2H2.  
 DR Pfam; PF00096; zf\_C2H2; 5.  
 DR SMART; SM00355; Znf\_C2H2; 5.  
 DR PROSITE; PS00815; AIPM\_HOMOCIT\_SYNTH\_1; 1.  
 DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 4.  
 DR PROSITE; PS00157; ZINC\_FINGER\_C2H2\_2; 5.  
 DR Metal-binding; Zinc; Zinc-finger.  
 KW Metal-binding; Zinc; Zinc-finger.  
 SQ SEQUENCE 2406 AA; 259381 MW; 7F498F2BFF3AB93F CRC64;

Query Match 90.5%; Score 38; DB 4; Length 2406;  
 Best Local Similarity 85.7%; Pred. No. 4; Se-02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
 Db 2316 PRPPRGR 2322

## RESULT 11

Q9HCL7 PRELIMINARY; PRT; 2414 AA.  
 ID Q9HCL7  
 AC Q9HCL7;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein KIAA1555 (Fragment).  
 GN KIAA1555  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=20450683; PubMed=10997877;  
 RA Nagase T., Kikuno R., Nakayama M., Hirotsawa M., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes.  
 RT xviii. The complete sequences of 100 new cDNA clones from brain which  
 RT code for large proteins in vitro.";  
 RL DNA Res. 7:273-281(2000).  
 DR EMBL; AB046775; BAB13381.2; -;  
 DR GO; GO:0016029; F:lyase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.

```

DR InterPro; IPR002034; AIPM/Hcit_synth.
DR InterPro; IPR007087; ZnF_C2H2.
DR Pfam; PF00096; ZF_C2H2_2.
DR SMART; SM00355; ZnF_C2H2_5.
DR PROSITE; PS00815; AIPM_HOMOCIT_SYNTH_1; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR PROSITE; PS01571; ZINC_FINGER_C2H2_2; 5.
DR KX Hypothetical protein.
DR NON_TER 1
FT SEQUENCE 2414 AA; 260294 MW; 05215E33CCEB2B65 CRC64;

Query Match 90.5%; Score 38; DB 4; Length 2414;
Best Local Similarity 85.7%; Pred.No. 4; 5e-02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7
DB 2324 PRPPQGR 2330

RESULT 12
Q86918 ID Q86918 PRELIMINARY; PRT; 72 AA.
AC Q86918;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Gag polyprotein (Fragment).
GN GAG.
OS Human T-lymphotropic virus 1.
OC Viruses; Retroid viruses; Retroviridae; Deltaretrovirus.
OX NCBI_TaxID=11908;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95306146; PubMed=7786587;
RA Major M., Baenke S., Nightingale S., Desselberger U.;
RT "Differential Tax expression in HTLV type 1-infected asymptomatic
carriers.";
RL AIDS Res. Hum. Retroviruses 11:415-421(1995).
DR EMBL; S78309; AAB34666.2; -.
DR EMBL; S78306; AAB34666.2; JOINED.
DR GO; GO:0019013; C: viral nucleocapsid; IEA.
DR GO; GO:0005196; F: structural molecule activity; IEA.
DR InterPro; IPR003139; Gag_p19.
DR Pfam; PF02228; Gag_p19; 1.
DR KX Polyprotein.
FT NON_TER 72
FT SEQUENCE 72 AA; 8039 MW; 1A4666800D766DA9 CRC64;

Query Match 88.1%; Score 37; DB 15; Length 72;
Best Local Similarity 100.0%; Pred.No. 25;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 6
DB 13 PRPPRGR 18

RESULT 13
Q9XZ63 ID Q9XZ63 PRELIMINARY; PRT; 117 AA.
AC Q9XZ63;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative secreted protein.
GN SCO3085 OR SCE25.26C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
[1]
RP SEQUENCE FROM N.A.

```

---

```

RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939115; CAB89457.1; -.
DR GO; GO:0016020; C: membrane; IEA.
DR GO; GO:0008519; F: ammonium transporter activity; IEA.
DR GO; GO:0006810; P: transport; IEA.
DR InterPro; IPR001905; Ammonium transpt.
DR Pfam; PF00909; Ammonium_transp; 1.
DR KX Complete proteome.
SQ SEQUENCE 117 AA; 13226 MW; 0C8F1151DD942D4C CRC64;

Query Match 88.1%; Score 37; DB 16; Length 117;
Best Local Similarity 85.7%; Pred.No. 40;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PRPPRGR 7
DB 6 PRPPRGR 12

RESULT 14
Q9XSS0 ID Q9XSS0 PRELIMINARY; PRT; 118 AA.
AC Q9XSS0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Fragment).
OS Canis familiaris (Dog).
OC Eukaryota; Euteleostomi; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thyroid;
RX MEDLINE=20422104; PubMed=10964405;
RA Pichon B., Mercan D., Pouillon V., Christophe-Hobertus C.,
RA Christophe D.;
RT targeting sequences on a functional basis.";
RL Anal. Biochem. 284:231-239(2000).
DR EMBL; AJ388548; CAB46847.1; -.

```

```
KW Hypothetical protein.
FT NON_TER 1 118
SQ SEQUENCE 118 AA; 1234 MW; A3DF7EC35749189C CRC64;

Query Match      88.1%; Score 37; DB 6; Length 118;
Best Local Similarity 100.0%; Pred.No. 40;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRG 6
DB 30 PRPPRG 35

RESULT 15
QYAM7
ID Q3YAM7 PRELIMINARY; PRT; 123 AA.
AC Q3YAM7;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein APL1916.
GN APL1916.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococccaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1."
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000062; BAA0921.1; -
DR PIR; D72579; D72579.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 123 AA; 1328 MW; 59BCA19F0191F468 CRC64;

Query Match      88.1%; Score 37; DB 17; Length 123;
Best Local Similarity 100.0%; Pred.No. 42;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRG 6
DB 31 PRPPRG 36

Search completed: April 6, 2004, 16:14:42
Job time : 25.3947 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:29:39 ; Search time 39.1869 Seconds  
(without alignments)  
50.472 Million cell updates/sec

Title: US-10-009-709-10

Perfect score: 42

Sequence: 1 PRPRGR 7

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : A\_Geneseq\_29Jan04:\*
- 1: Geneseqp1980s:\*
  - 2: Geneseqp1990s:\*
  - 3: Geneseqp2000s:\*
  - 4: Geneseqp2001s:\*
  - 5: Geneseqp2002s:\*
  - 6: Geneseqp2003as:\*
  - 7: Geneseqp2003bs:\*
  - 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	42	100.0	7	4 AAB48780	Aab48780 Human sal
2	42	100.0	8	4 AAB48779	Aab48779 Human sal
3	42	100.0	9	4 AAB48778	Aab48778 Human sal
4	42	100.0	10	4 AAB48777	Aab48777 Human sal
5	42	100.0	17	4 AAB48783	Aab48783 Human sal
6	42	100.0	92	6 ABM65798	Abm65798 Propionib
7	42	100.0	104	4 ABG18278	Abg18278 Novel hum
8	42	100.0	132	4 ABG38848	Abg38848 Peptide #
9	42	100.0	132	4 AAM32323	Aam32323 Peptide #
10	42	100.0	132	4 AAM72058	Aam72058 Human bon
11	42	100.0	132	4 AAM59494	Aam59494 Human bra
12	42	100.0	132	4 ABG53742	Abg53742 Human liv
13	42	100.0	132	5 ABG41873	Abg41873 Human pep
14	42	100.0	141	4 AAU65747	Aau65747 Propionib
15	42	100.0	141	6 ABM62266	Abm62266 Propionib
16	42	100.0	149	6 ABR57423	Abm57423 Human NOV
17	42	100.0	154	6 ABR57423	Abm57423 Human sec
18	42	100.0	166	6 ADA83798	Ada83798 Human PRH
19	42	100.0	166	7 ADC98216	Adc98216 Human sal
20	42	100.0	258	4 ABG26281	Abg26281 Novel hum
21	42	100.0	306	4 ABG03818	Abg03818 Novel hum
22	42	100.0	578	7 ADE07941	Ade07941 Novel pro
23	42	100.0	694	5 ABG70853	Abg70853 Streptoco
24	42	100.0	1938	6 ABP76681	Abp76681 Streptomy
25	39	92.9	91	4 AAO01213	Aao01213 Human pol

26	39	92.9	290	4	AAO05792	Aao05792 Human pol
27	39	92.9	547	7	ADC37585	Adc37585 Human nuc
28	38	90.5	74	4	AAU40822	Aau40822 Propionib
29	38	90.5	74	6	ABM37341	Abm37341 Propionib
30	38	90.5	79	4	AAU40575	Aau40575 Propionib
31	38	90.5	79	6	ABM37094	Abm37094 Propionib
32	38	90.5	1708	5	ABP69375	Abp69375 Human pol
33	38	90.5	2405	6	ABB2733	Abb2733 Human KRC
34	37	88.1	17	2	AAU99907	Aau99907 Peptide e
35	37	88.1	46	3	AAU99907	Aau99907 Peptide e
36	37	88.1	46	4	AAU19074	Aau19074 Peptide #
37	37	88.1	46	4	ABB38280	Abb38280 Peptide #
38	37	88.1	46	4	AAU31712	Aau31712 Peptide #
39	37	88.1	46	4	ABB23461	Abb23461 Protein #
40	37	88.1	46	4	AAU71425	Aau71425 Human bon
41	37	88.1	46	4	AAU58902	Aau58902 Human bra
42	37	88.1	46	4	ABG53123	Abg53123 Human liv
43	37	88.1	46	5	ABG41232	Abg41232 Human pep
44	37	88.1	56	4	AAU67167	Aau67167 Propionib
45	37	88.1	56	6	ABM63686	Abm63686 Propionib

ALIGNMENTS

RESULT 1  
AAB48780  
ID AAB48780 standard; peptide; 7 AA.  
XX  
AC AAB48780;  
XX  
DT 09-MAR-2001 (first entry)  
XX  
DE Human saliva PRP-1 fragment (residues 102-108), SEQ ID NO:10.  
XX  
KW Human; PRP-1; proline-rich protein; saliva; dental caries;  
KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
KW oral bacterium; caries prevention.  
XX  
OS Homo sapiens.  
XX  
PN WO2000069890-A1.  
XX  
PD 23-NOV-2000.  
XX  
PF 11-MAY-2000; 2000WO-SE000930.  
XX  
PR 17-MAY-1999; 99SE-00001773.  
XX  
PA (STRO/) STROEMBERG N.  
PA (JOHA/) JOHANSSON I.  
XX  
PI Stroemberg N, Johansson I;  
XX  
DR WPI; 2001-031923/04.  
XX  
PT New oligopeptides comprising 2 arginine residues from degradation of  
PT proline-rich proteins, useful for preventing dental caries.  
XX  
PS Claim 4; Page 24; 36pp; English.  
XX  
CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
CC AAB48783) which contain at least two arginine residues and which protect  
CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
CC determinants of a person's susceptibility to dental caries. PRPs are  
CC degraded by Actinomyces and Streptococcus species to small peptide  
CC fragments, these are metabolised by oral bacteria for nutritional  
CC purposes, with certain bacterial species generating ammonia via the  
CC catabolism of arginine. The peptides of the invention, being arginine-  
CC rich, can also be converted to ammonia by these bacteria. The ammonia  
CC thus formed raises the pH at the dental surface, thereby protecting the  
CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-

CC derived oligopeptides of the invention

XX Sequence 7 AA;

XX Query Match 100.0%; Score 42; DB 4; Length 7;  
XX Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
SQ Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7

Db 1 PRPPRGR 7

RESULT 2

XX AAB48779 standard; peptide; 8 AA.

XX AC AAB48779;

XX DT 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 101-108), SEQ ID NO:9.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;  
XX chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
XX oral bacterium; caries prevention.

XX OS Homo sapiens.

XX WO200069890-A1.

XX PD 23-NOV-2000.

XX PF 11-MAY-2000; 2000WO-SE000930.

XX PR 17-MAY-1999; 99SE-00001773.

XX (STRO/) STROEMBERG N.

XX (JOHA/) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of  
XX proline-rich proteins, useful for preventing dental caries.

XX Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
XX AAB48783) which contain at least two arginine residues and which protect  
XX against dental caries. PRPs (proline-rich proteins) are salivary proteins  
XX encoded by six clustered genes on chromosome 12p13.2 and are potential  
XX determinants of a person's susceptibility to dental caries. PRPs are  
XX degraded by Actinomyces and Streptococcus species to small peptide  
XX fragments. These are metabolised by oral bacteria for nutritional  
XX catabolism of arginine. The peptides of the invention, being arginine-  
XX rich, can also be converted to ammonia by these bacteria. The ammonia  
XX thus formed raises the pH at the dental surface, thereby protecting the  
XX teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
XX derived oligopeptides of the invention

XX Sequence 8 AA;

XX Query Match 100.0%; Score 42; DB 4; Length 8;  
XX Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7

Db 2 PRPPRGR 8

RESULT 3

XX AAB48778

XX ID AAB48778 standard; peptide; 9 AA.

XX AC AAB48778;

XX DT 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 100-108), SEQ ID NO:8.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;  
XX chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
XX oral bacterium; caries prevention.

XX OS Homo sapiens.

XX WO200069890-A1.

XX PD 23-NOV-2000.

XX PF 11-MAY-2000; 2000WO-SE000930.

XX PR 17-MAY-1999; 99SE-00001773.

XX (STRO/) STROEMBERG N.

XX (JOHA/) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of  
XX proline-rich proteins, useful for preventing dental caries.

XX Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
XX AAB48783) which contain at least two arginine residues and which protect  
XX against dental caries. PRPs (proline-rich proteins) are salivary proteins  
XX encoded by six clustered genes on chromosome 12p13.2 and are potential  
XX determinants of a person's susceptibility to dental caries. PRPs are  
XX degraded by Actinomyces and Streptococcus species to small peptide  
XX fragments. These are metabolised by oral bacteria for nutritional  
XX purposes, with certain bacterial species generating ammonia via the  
XX catabolism of arginine. The peptides of the invention, being arginine-  
XX rich, can also be converted to ammonia by these bacteria. The ammonia  
XX thus formed raises the pH at the dental surface, thereby protecting the  
XX teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
XX derived oligopeptides of the invention

XX Sequence 9 AA;

XX Query Match 100.0%; Score 42; DB 4; Length 9;  
XX Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7

Db 3 PRPPRGR 9

RESULT 4

XX AAB48777

XX ID AAB48777 standard; peptide; 10 AA.

XX AC AAB48777;

XX DT 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 99-108), SEQ ID NO:7.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
KW oral bacterium; caries prevention.  
XX Homo sapiens.  
XX WO200069890-A1.  
XX 23-NOV-2000.  
XX 11-MAY-2000; 2000WO-SE000930.  
XX 17-MAY-1999; 99SE-00001773.  
XX (STRO/) STROEMBERG N.  
XX (JOHA/) JOHANSSON I.  
XX Stroemberg N, Johansson I;  
XX WPI; 2001-031923/04.  
XX New oligopeptides comprising 2 arginine residues from degradation of  
XX proline-rich proteins, useful for preventing dental caries.  
XX Claim 4; Page 24; 36pp; English.  
XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
XX AAB48783) which contain at least two arginine residues and which protect  
XX against dental caries. PRPs (proline-rich proteins) are salivary proteins  
XX encoded by six clustered genes on chromosome 12p13.2 and are potential  
XX determinants of a person's susceptibility to dental caries. PRPs are  
XX degraded by Actinomyces and Streptococcus species to small peptide  
XX fragments. These are metabolised by oral bacteria for nutritional  
XX purposes, with certain bacterial species generating ammonia via the  
XX catabolism of arginine. The peptides of the invention, being arginine-  
XX rich, can also be converted to ammonia by these bacteria. The ammonia  
XX thus formed raises the pH at the dental surface, thereby protecting the  
XX teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
XX derived oligopeptides of the invention  
XX Sequence 10 AA;  
XX Query Match 100.0%; Score 42; DB 4; Length 10;  
XX Best Local Similarity 100.0%; Pred. No. 5.6;  
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PRPPRGR 7  
DB |||||  
4 PRPPRGR 10  
RESULT 5  
AAB48783  
ID AAB48783 standard; peptide; 17 AA.  
XX  
XX AAB48783;  
XX  
XX 09-MAR-2001 (first entry)  
XX Human saliva PRP-1 fragment (residues 99-115), SEQ ID NO:13.  
XX  
XX Human; PRP-1; proline-rich protein; saliva; dental caries;  
KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
KW oral bacterium; caries prevention.  
XX Homo sapiens.  
XX WO200069890-A1.  
XX 23-NOV-2000.  
XX 11-MAY-2000; 2000WO-SE000930.  
XX 17-MAY-1999; 99SE-00001773.

XX (STRO/) STROEMBERG N.  
XX (JOHA/) JOHANSSON I.  
XX Stroemberg N, Johansson I;  
XX WPI; 2001-031923/04.  
XX New oligopeptides comprising 2 arginine residues from degradation of  
XX proline-rich proteins, useful for preventing dental caries.  
XX Claim 2; Page 24; 36pp; English.  
XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
XX AAB48783) which contain at least two arginine residues and which protect  
XX against dental caries. PRPs (proline-rich proteins) are salivary proteins  
XX encoded by six clustered genes on chromosome 12p13.2 and are potential  
XX determinants of a person's susceptibility to dental caries. PRPs are  
XX degraded by Actinomyces and Streptococcus species to small peptide  
XX fragments. These are metabolised by oral bacteria for nutritional  
XX purposes, with certain bacterial species generating ammonia via the  
XX catabolism of arginine. The peptides of the invention, being arginine-  
XX rich, can also be converted to ammonia by these bacteria. The ammonia  
XX thus formed raises the pH at the dental surface, thereby protecting the  
XX teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
XX derived oligopeptides of the invention  
XX Sequence 17 AA;  
XX Query Match 100.0%; Score 42; DB 4; Length 17;  
XX Best Local Similarity 100.0%; Pred. No. 8.8;  
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PRPPRGR 7  
DB |||||  
4 PRPPRGR 10  
RESULT 6  
AAB48778  
ID AAB48778 standard; protein; 92 AA.  
XX  
XX AAB48778;  
XX  
XX 20-OCT-2003 (first entry)  
XX  
XX Propionibacterium acnes immunogenic polypeptide #30474.  
XX  
XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
KW immunostimulant; immune response; vaccine; immunogenic.  
XX  
XX Propionibacterium acnes.  
XX  
XX WO2003033515-A1.  
XX  
XX 24-APR-2003.  
XX  
XX 11-OCT-2002; 2002WO-US032727.  
XX  
XX 15-OCT-2001; 2001US-00978825.  
XX  
XX (CORI-) CORIXA CORP.  
XX  
XX Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
PI Barth B, Vallieve-Douglass J;  
XX WPI; 2003-381789/36.  
XX  
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
XX polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
XX or for stimulating an immune response specific for a P. acnes protein.

PS Claim 7; SEQ ID NO 30474; 1481bp; English.  
XX  
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
CC encoding a Propionibacterium acnes protein. The invention also relates to  
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
CC immunogenic fragments of P. acnes polypeptides. The invention  
CC additionally encompasses expression vectors and host cells comprising a  
CC polynucleotide of the invention; antibodies against polypeptides of the  
CC invention; fusion proteins comprising a polypeptide of the invention; a  
CC method for stimulating an immune response specific for a P. acnes  
CC polypeptide and an isolated T cell population comprising T cells prepared  
CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
CC antigen-presenting cells that express the polypeptide); a method and kit  
CC for detecting or determining the presence or absence of P. acnes in a  
CC patient; and a method for inhibiting the development of P. acnes in a  
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
CC proteins, T cell populations or antigen-presenting cells that express the  
CC polypeptides are useful for diagnosing, preventing or treating acne  
CC vulgaris, or for stimulating an immune response specific for a P. acnes  
CC protein. The polynucleotides can also be used as probes or primers for  
CC nucleic acid hybridisation. The vaccine composition is useful for the  
CC stimulation of an immune response against P. acnes, or for treating acne,  
CC and the kit is useful for performing a diagnostic assay. The present  
CC sequence represents a specifically claimed P. acnes polypeptide which is  
CC thought to contain an immunogenic region. Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 92 AA;

Query Match 100.0%; Score 42; DB 6; Length 92;  
Best Local Similarity 100.0%; Pred. No. 37;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPPRGR 7  
Db 17 PRPPRGR 23  
|||||

RESULT 7  
ABG18278  
ID ABG18278 standard; protein; 104 AA.

AC ABG18278;  
XX  
DT 18-FEB-2002 (first entry)  
DE Novel human diagnostic protein #18269.  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX WO200175067-A2.  
PN 11-OCT-2001.  
PD  
XX 30-MAR-2001; 2001WO-US008631.  
PF  
XX 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
PR (HYSE-) HYSEQ INC.  
XX  
PA Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS82465.  
DR  
XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX

PS Claim 20; SEQ ID NO 48637; 103pp; English.

XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX

SQ Sequence 104 AA;

Query Match 100.0%; Score 42; DB 4; Length 104;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPPRGR 7  
Db 93 PRPPRGR 99  
|||||

RESULT 8  
ABB38848  
ID ABB38848 standard; peptide; 132 AA.

AC ABB38848;  
XX  
DT 04-FEB-2002 (first entry)  
DE Peptide #6354 encoded by human foetal liver single exon probe.  
XX Human; foetal liver; gene expression; single exon nucleic acid probe.  
XX Homo sapiens.  
XX WO200157277-A2.  
PN 09-AUG-2001.  
PD  
XX 30-JAN-2001; 2001WO-US000669.  
PF  
XX 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GE-00024263.  
XX (MOLE-) MOLECULAR DYNAMICS INC.  
XX Penn SG, Hanzel DK, Chen W, Rank DR;  
PI WPI; 2001-483447/52.  
DR  
XX



PT Human genome-derived single exon nucleic acid probes useful for analyzing  
 PT gene expression in human fetal liver.

PS Claim 27; SEQ ID NO 31483; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring  
 CC human gene expression in a sample derived from human foetal liver. The  
 CC single exon nucleic acid probes may be used for predicting, measuring and  
 CC displaying gene expression in samples derived from human foetal liver. The  
 CC present sequence is a peptide encoded by a single exon nucleic acid probe  
 CC of the invention. Note: The sequence data for this patent did not form  
 CC part of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 132 AA;

SQ Query Match 100.0%; Score 42; DB 4; Length 132;

Best Local Similarity 100.0%; Pred. No. 51;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPRGR 7

Db 84 PRPRGR 90

RESULT 9

AAAM32323  
 ID AAAM32323 standard; protein; 132 AA.

XX AC AAAM32323;

XX 17-OCT-2001 (first entry)

XX Peptide #6360 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder.

XX Homo sapiens.

XX WO200157272-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000663.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488997/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human placenta.

XX Claim 27; SEQ ID NO 32592; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP:  
 CC see AA131315-AA157546). The present sequence is a peptide encoded by one  
 CC such probe. The probes are useful for producing a microarray for  
 CC predicting, measuring and displaying gene expression in samples derived  
 CC from human placenta. The probes are useful for antenatal diagnosis of  
 CC human genetic disorders

XX Sequence 132 AA;

Query Match 100.0%; Score 42; DB 4; Length 132;

Best Local Similarity 100.0%; Pred. No. 51;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPRGR 7

Db 84 PRPRGR 90

RESULT 10

AAAM72058

ID AAAM72058 standard; protein; 132 AA.

XX AC AAAM72058;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed probe encoded protein SEQ ID NO: 32364.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

XX microarray; cancer; leukaemia; lymphoma; myeloma.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US000668.

XX 04-FEB-2000; 2000US-0180312P.

XX 26-MAY-2000; 2000US-0207456P.

XX 30-JUN-2000; 2000US-00608408.

XX 03-AUG-2000; 2000US-00632366.

XX 21-SEP-2000; 2000US-0234687P.

XX 27-SEP-2000; 2000US-0236359P.

XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488990/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human bone marrow.

XX Example 4; SEQ ID NO 32364; 658pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid  
 CC probes which are derived from genomic sequences expressed in the human  
 CC bone marrow. They can be used to measure gene expression in bone marrow  
 CC samples, which may enable the improved diagnosis and treatment of cancers  
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a  
 CC protein encoded by one of the probes of the invention

XX Sequence 132 AA;

Query Match 100.0%; Score 42; DB 4; Length 132;

Best Local Similarity 100.0%; Pred. No. 51;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PRPRGR 7

Db 84 PRPRGR 90

RESULT 11

AAAM59494

ID AAAM59494 standard; protein; 132 AA.

XX AC AAAM59494;

```

XX 05-NOV-2001 (first entry)
XX Human brain expressed single exon probe encoded protein SEQ ID NO: 31599.
XX Human; brain expressed exon; gene expression analysis; probe; microarray;
XX Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX Homo sapiens.
XX WO200157275-A2.
XX 09-AUG-2001.
XX 30-JAN-2001; 2001WO-US000667.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX Example 4; SEQ ID NO 31599; 650pp + Sequence Listing; English.
XX The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention
XX Sequence 132 AA;
XX Query Match 100.0%; Score 42; DB 4; Length 132;
XX Best Local Similarity 100.0%; Pred. NO. 51;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX Qy 1 PRPRGR 7
XX Db 84 PRPRGR 90
XX
XX RESULT 12
XX AEG53742
XX ID AEG53742 standard; peptide; 132 AA.
XX AC AEG53742;
XX 25-FEB-2003 (first entry)
XX Human liver peptide, SEQ ID NO 32390.
XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX hypercholesterolaemia; coronary heart disease.
XX Homo sapiens.
XX WO200157273-A2.
XX 09-AUG-2001.

```

```

XX 30-JAN-2001; 2001WO-US000664.
XX 04-FEB-2000; 2000US-0180312P.
XX 26-MAY-2000; 2000US-0207456P.
XX 30-JUN-2000; 2000US-00608408.
XX 03-AUG-2000; 2000US-00632366.
XX 21-SEP-2000; 2000US-0234687P.
XX 27-SEP-2000; 2000US-0236359P.
XX 04-OCT-2000; 2000GB-00024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488899/53.
XX Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX Claim 27; SEQ ID NO 32390; 658pp; English.
XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridizes at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.
XX (I) may be used for predicting, measuring and displaying gene expression
XX in samples derived from human adult liver. The genes identified may be
XX involved in genetic liver diseases such as cirrhosis,
XX hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
XX associated with coronary heart disease. AEG47348-ABG59930 represent human
XX liver single exon encoded peptides of the invention. Note: The sequence
XX information for this patent does not appear in the printed specification
XX but was obtained in electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 132 AA;
XX Query Match 100.0%; Score 42; DB 4; Length 132;
XX Best Local Similarity 100.0%; Pred. NO. 51;
XX Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX Qy 1 PRPRGR 7
XX Db 84 PRPRGR 90
XX
XX RESULT 13
XX AEG41873
XX ID AEG41873 standard; peptide; 132 AA.
XX AC AEG41873;
XX 19-AUG-2002 (first entry)
XX Human peptide encoded by genome-derived single exon probe SEQ ID 31538.
XX Human; single exon probe; asthma; lung cancer; COPD; IiD;
XX chronic obstructive pulmonary disease; interstitial lung disease;
XX familial idiopathic pulmonary fibrosis; neurofibromatosis;
XX tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;
XX Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemorrhoidosis;
XX pulmonary histiocytosis; lymphangioleiomyomatosis; Karagener syndrome;
XX pulmonary alveolar proteinosis; fibrocystic pulmonary dysplasia;
XX primary ciliary dyskinesia; pulmonary hypertension;
XX hyaline membrane disease.
XX Homo sapiens.
XX WO200186003-A2.
XX 15-NOV-2001.

```

PF 30-JAN-2001; 2001WO-US000665.  
XX  
XX 04-FEB-2000; 2000US-0180312P.  
PR 26-MAY-2000; 2000US-0207456P.  
PR 30-JUN-2000; 2000US-00608408.  
PR 03-AUG-2000; 2000US-00632366.  
PR 21-SEP-2000; 2000US-0234687P.  
PR 27-SEP-2000; 2000US-0236359P.  
PR 04-OCT-2000; 2000GB-00024263.  
XX  
XX (MOLE-) MOLECULAR DYNAMICS INC.

PI Penn SG, Hanzel DK, Chen W, Rank DR;  
XX  
XX WPI; 2002-114163/15.

PT Spatially-addressable set of single exon nucleic acid probes, used to  
PT measure gene expression in human lung samples.

XX  
XX Claim 27; SEQ ID NO 31538; 634pp; English.

XX The invention relates to a spatially-addressable set of single exon  
CC nucleic acid probes for measuring gene expression in a sample derived  
CC from human lung comprising single exon nucleic acid probes having one of  
CC 12614 nucleic acid sequences mentioned in the specification, or their  
CC complements or the 12387 open reading frames derived from the 12614  
CC probes. Also included are a microarray comprising the novel set of probes  
CC; the novel set of probes which hybridise at high stringency to a nucleic  
CC acid expressed in the human lung; measuring gene expression in a sample  
CC derived from human lung, comprising (a) contacting the array with a  
CC collection of detectably labeled nucleic acids derived from human lung  
CC mRNA, and (b) measuring the label detectably bound to each probe of the  
CC array; identifying exons in a eukaryotic genome, comprising (a)  
CC algorithmically predicting at least one exon from genomic sequences of  
CC the eukaryote; and (b) detecting specific hybridisation of detectably  
CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
CC having a fragment identical to the predicted exon, the probe is included  
CC in the above mentioned microarray; assigning exons to a single gene,  
CC comprising (a) identifying exons from genomic sequence by the method  
CC above and (b) measuring the expression of each of the exons in several  
CC tissues and/or cell types using hybridisation to a single exon  
CC microarrays having a probe with the exon, where a common pattern of  
CC expression of the exons in the tissues and/or cell types indicates that  
CC the exons should be assigned to a single gene; a peptide comprising one  
CC of 12011 sequences, mentioned in the specification, or encoded by the  
CC probes/open reading frames (ORF). The probes are used for gene expression  
CC analysis, and for identifying exons in a gene, particularly using human  
CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
CC histiocytosis, lymphangioleiomyomatosis, pulmonary alveolar proteinosis,  
CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
CC present sequence is a peptide/protein encoded by a single exon probe of  
CC the invention. Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 42; DB 5; Length 132;  
Best Local Similarity 100.0%; Pred. No. 51;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
Db |||||  
84 PRPPRGR 90

RESULT 14  
AAU65747

ID AAU65747 standard; protein; 141 AA.  
XX  
XX AC AAU65747;  
XX  
XX 27-FEB-2002 (first entry)  
XX  
XX Propionibacterium acnes immunogenic protein #26643.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
KW dermatological; osteopathic; neuroprotectant.

OS Propionibacterium acnes.

PN WO200181581-A2.

XX  
XX 01-NOV-2001.

XX  
XX 20-APR-2001; 2001WO-US013865.

XX  
XX 21-APR-2000; 2000US-0199047P.

PR 02-JUN-2000; 2000US-0208941P.

PR 07-JUL-2000; 2000US-0216747P.

XX  
XX (CORI-) CORIXA CORP.

PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX  
XX WPI; 2001-616774/71.

DR N-PSDB; AAS59689.

XX Propionibacterium acnes polypeptides and nucleic acids useful for  
PT vaccinating against and diagnosing infections, especially useful for  
PT treating acne vulgaris.

XX Example 1; SEQ ID NO 26942; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
CC P. acnes is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 141 AA;

Query Match 100.0%; Score 42; DB 4; Length 141;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
Db |||||  
17 PRPPRGR 23

RESULT 15  
ABM62266

ID XX ABM62266 standard; protein; 141 AA.  
XX AC ABM62266;  
XX DT 20-OCT-2003 (first entry)  
XX DE Propionibacterium acnes predicted ORF-encoded polypeptide #26942.  
XX KW Acne vulgaris; antisborrheic; dermatological; antibacterial;  
XX KW immunostimulant; immune response; vaccine.  
XX OS Propionibacterium acnes.  
XX PN WO2003033515-A1.  
XX PD 24-APR-2003.  
XX PF 11-OCT-2002; 2002WO-US032727.  
XX PR 15-OCT-2001; 2001US-00978825.  
XX PA (CORI-) CORIXA CORP.  
XX PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
PI Zhang Y, Wang S, Jen S, Lodes WJ, Benson DR, Jones R, Carter D;  
PI Barth B, Valliave-Douglass J;  
XX WI: 2003-381789/36.  
DR N-PSDB; ACF64618.  
XX  
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
PT or for stimulating an immune response specific for a P. acnes protein.  
XX  
XX Example 1; SEQ ID NO 26942; 1481pp; English.  
XX  
XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
CC encoding a Propionibacterium acnes protein. The invention also relates to  
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
CC immunogenic fragments of P. acnes polypeptides. The invention  
CC additionally encompasses expression vectors and host cells comprising a  
CC polynucleotide of the invention; antibodies against polypeptides of the  
CC invention; fusion proteins comprising a polypeptide of the invention; a  
CC method for stimulating an immune response specific for a P. acnes  
CC polypeptide and an isolated T cell population comprising P. acnes polypeptides,  
CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
CC antigen-presenting cells that express the polypeptide); a method and kit  
CC for detecting or determining the presence or absence of P. acnes in a  
CC patient; and a method for inhibiting the development of P. acnes in a  
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
CC proteins, T cell populations or antigen-presenting cells that express the  
CC polypeptides are useful for diagnosing, preventing or treating acne  
CC vulgaris, or for stimulating an immune response specific for a P. acnes  
CC protein. The polynucleotides can also be used as probes or primers for  
CC nucleic acid hybridisation. The vaccine composition is useful for the  
CC stimulation of an immune response against P. acnes, or for treating acne,  
CC and the kit is useful for performing a diagnostic assay. The present  
CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
CC reading frame) contained within the P. acnes polynucleotides of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 141 AA;  
SQ

Query Match 100.0%; Score 42; DB 6; Length 141;  
Best Local Similarity 100.0%; Pred. No. 54;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
17 PRPPRGR 23

Search completed: April 6, 2004, 16:06:41  
Job time : 40.1869 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 16:14:50 ; Search time 27.4766 Seconds  
(without alignments)  
66.909 Million cell updates/sec

Title: US-10-009-709-10

Perfect score: 42

Sequence: 1 PRPGR 7

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 26263353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications AA.\*

- 1: /cgn2\_6/prodata/1/pubpaa/PCT\_NEW\_PUB.pdb.\*
- 2: /cgn2\_6/prodata/1/pubpaa/PCT\_NEW\_PUB.pdb.\*
- 3: /cgn2\_6/prodata/1/pubpaa/US05\_NEW\_PUB.pdb.\*
- 4: /cgn2\_6/prodata/1/pubpaa/US05\_PUBCOMB.pdb.\*
- 5: /cgn2\_6/prodata/1/pubpaa/US07\_NEW\_PUB.pdb.\*
- 6: /cgn2\_6/prodata/1/pubpaa/PCTUS\_PUBCOMB.pdb.\*
- 7: /cgn2\_6/prodata/1/pubpaa/US08\_NEW\_PUB.pdb.\*
- 8: /cgn2\_6/prodata/1/pubpaa/US08\_PUBCOMB.pdb.\*
- 9: /cgn2\_6/prodata/1/pubpaa/US09A\_PUBCOMB.pdb.\*
- 10: /cgn2\_6/prodata/1/pubpaa/US09B\_PUBCOMB.pdb.\*
- 11: /cgn2\_6/prodata/1/pubpaa/US09C\_PUBCOMB.pdb.\*
- 12: /cgn2\_6/prodata/1/pubpaa/US09D\_PUBCOMB.pdb.\*
- 13: /cgn2\_6/prodata/1/pubpaa/US10A\_PUBCOMB.pdb.\*
- 14: /cgn2\_6/prodata/1/pubpaa/US10B\_PUBCOMB.pdb.\*
- 15: /cgn2\_6/prodata/1/pubpaa/US10C\_PUBCOMB.pdb.\*
- 16: /cgn2\_6/prodata/1/pubpaa/US10\_NEW\_PUB.pdb.\*
- 17: /cgn2\_6/prodata/1/pubpaa/US60\_NEW\_PUB.pdb.\*
- 18: /cgn2\_6/prodata/1/pubpaa/US60\_PUBCOMB.pdb.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	72	12	US-10-424-599-223703
2	42	100.0	132	9	Sequence 223703, Sequence 43644, A
3	42	100.0	156	14	Sequence 80, Appl
4	42	100.0	1962	15	Sequence 7, Appl
5	39	92.9	71	12	Sequence 178111, Sequence 285441, A
6	38	90.5	94	12	Sequence 285441, A
7	37	88.1	40	14	Sequence 106, App
8	37	88.1	40	15	Sequence 106, App
9	37	88.1	46	9	Sequence 38759, A
10	37	88.1	71	12	Sequence 151579, A
11	37	88.1	95	9	Sequence 1824, App
12	37	88.1	96	12	Sequence 1824, App
13	37	88.1	131	12	Sequence 151513, A
14	37	88.1	209	15	Sequence 52873, A
15	37	88.1	219	16	Sequence 1429, App

15	37	88.1	238	12	US-10-425-114-69167
17	37	88.1	247	15	Sequence 69167, A
18	37	88.1	262	12	Sequence 2734, App
19	37	88.1	277	12	Sequence 59553, A
20	37	88.1	291	12	Sequence 43478, A
21	37	88.1	401	15	Sequence 43478, A
22	37	88.1	429	9	Sequence 60421, A
23	37	88.1	440	14	Sequence 5026, App
24	37	88.1	440	14	Sequence 6, Appl
25	37	88.1	674	11	Sequence 66, Appl
26	37	88.1	1471	8	Sequence 70, Appl
27	37	88.1	1474	14	Sequence 69, Appl
28	37	88.1	1474	15	Sequence 1, Appl
29	37	88.1	1515	14	Sequence 522, App
30	36	85.7	62	9	Sequence 914, App
31	36	85.7	75	12	Sequence 8, Appl
32	36	85.7	213	12	Sequence 47, Appl
33	36	85.7	272	12	Sequence 183321, A
34	36	85.7	373	15	Sequence 248978, A
35	36	85.7	377	14	Sequence 41709, A
36	36	85.7	447	10	Sequence 3087, App
37	36	85.7	447	14	Sequence 10728, A
38	36	85.7	447	14	Sequence 10, Appl
39	36	85.7	447	15	Sequence 132, App
40	36	85.7	447	15	Sequence 78, Appl
41	36	85.7	448	14	Sequence 112, App
42	36	85.7	448	14	Sequence 132, App
43	36	85.7	448	14	Sequence 3, Appl
44	36	85.7	449	14	Sequence 1, Appl
45	36	85.7	449	14	Sequence 1, Appl

#### ALIGNMENTS

##### RESULT 1

US-10-424-599-223703  
; Sequence 223703, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 223703  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_44033C.1.pep  
US-10-424-599-223703

Query Match 100.0%; Score 42; DB 12; Length 72;  
Best Local Similarity 100.0%; Pred.No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPGR 7

Db 27 PRPGR 33

##### RESULT 2

US-09-864-761-43644  
; Sequence 43644, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharron G.

```
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wenheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ FILE REFERENCE: Aomic-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ PRIOR FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 43644
/ LENGTH: 132
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC006518.17
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77
/ OTHER INFORMATION: EST HUMAN HIT: BF089785.1, EVALUE 1.00e-01
/ OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02
/ US-09-864-761-43644
Query Match      100.0%; Score 42; DB 9; Length 132;
Best Local Similarity 100.0%; Pred. No. 68;
Matches       7; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

QY      1 PRPPRGR 7
DB      84 FRPFRGR 90

RESULT 3
US-10-157-031-80

/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wenheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ FILE REFERENCE: Aomic-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ PRIOR FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 43644
/ LENGTH: 132
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC006518.17
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77
/ OTHER INFORMATION: EST HUMAN HIT: BF089785.1, EVALUE 1.00e-01
/ OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02
/ US-09-864-761-43644
Query Match      100.0%; Score 42; DB 9; Length 132;
Best Local Similarity 100.0%; Pred. No. 68;
Matches       7; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

QY      1 PRPPRGR 7
DB      84 FRPFRGR 90

RESULT 3
US-10-157-031-80

/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wenheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ FILE REFERENCE: Aomic-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ PRIOR FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04
/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-08-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 43644
/ LENGTH: 132
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AC006518.17
/ OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
/ OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
/ OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77
/ OTHER INFORMATION: EST HUMAN HIT: BF089785.1, EVALUE 1.00e-01
/ OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02
/ US-09-864-761-43644
Query Match      100.0%; Score 42; DB 9; Length 132;
Best Local Similarity 100.0%; Pred. No. 68;
Matches       7; Conservative    0; Mismatches    0; Indels    0; Gaps    0;

QY      1 PRPPRGR 7
DB      84 FRPFRGR 90

RESULT 3
US-10-157-031-80
```

; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 178111  
; LENGTH: 71  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_131850C.1.pap  
US-10-424-599-178111

Query Match 92.9%; Score 39; DB 12; Length 71;  
Best Local Similarity 85.7%; Pred. No. 1.1e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
DB 61 PRPPRGR 67

RESULT 6  
US-10-424-599-285441  
; Sequence 285441, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 285441  
; LENGTH: 94  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_99782C.1.pap  
US-10-424-599-285441

Query Match 90.5%; Score 38; DB 12; Length 94;  
Best Local Similarity 85.7%; Pred. No. 1.9e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
DB 41 PRPPRGR 47

RESULT 7  
US-10-008-524A-106  
; Sequence 106, Application US/10008524A  
; Publication No. US20030175682A1  
; GENERAL INFORMATION:  
; APPLICANT: Doorbar, John  
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA  
; TITLE OF INVENTION: VIRUSES  
; FILE REFERENCE: 18396/1074  
; CURRENT APPLICATION NUMBER: US/10/008,524A  
; CURRENT FILING DATE: 2002-08-13  
; PRIOR APPLICATION NUMBER: 09/314,268  
; PRIOR FILING DATE: 1999-05-18  
; NUMBER OF SEQ ID NOS: 179  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 106

; LENGTH: 40  
; TYPE: PRT  
; ORGANISM: Human papillomavirus type 30  
US-10-008-524A-106

Query Match 88.1%; Score 37; DB 14; Length 40;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
DB 10 PRPPRGR 16

RESULT 8  
US-10-350-719-106  
; Sequence 106, Application US/10350719  
; Publication No. US20030219728A1  
; GENERAL INFORMATION:  
; APPLICANT: Doorbar, John  
; TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA  
; TITLE OF INVENTION: VIRUSES  
; FILE REFERENCE: 18396/2162  
; CURRENT APPLICATION NUMBER: US/10/350,719  
; CURRENT FILING DATE: 2003-01-24  
; PRIOR APPLICATION NUMBER: PCT/GB01/01176  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: GB0018140.4  
; PRIOR FILING DATE: 2000-07-24  
; NUMBER OF SEQ ID NOS: 180  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 106  
; LENGTH: 40  
; TYPE: PRT  
; ORGANISM: Human papillomavirus type 30  
US-10-350-719-106

Query Match 88.1%; Score 37; DB 15; Length 40;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
DB 10 PRPPRGR 16

RESULT 9  
US-09-864-761-38759  
; Sequence 38759, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
; FILE REFERENCE: Aecmics-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.6  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/00666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/00667

PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 38759  
LENGTH: 46  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AB016897.1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.9  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 4.9  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.6  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.7  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.7  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.5  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 6.5  
US-09-864-761-38759

Query Match 88.1%, Score 37; DB 9; Length 46;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRG 6  
Db 8 PRPPRG 13

RESULT 10  
US-10-424-599-151579  
Sequence 151579, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 151579  
LENGTH: 71  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_10789C.1.pep  
US-10-424-599-151579

Query Match 88.1%, Score 37; DB 12; Length 71;  
Best Local Similarity 100.0%; Pred. No. 2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRG 6  
Db 66 PRPPRG 71

RESULT 11  
US-09-925-300-1824  
Sequence 1824, Application US/09925300  
Patent No. US20020151681A1  
GENERAL INFORMATION:  
APPLICANT: Craig Ruben  
APPLICANT: Steve Ruben  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA101  
CURRENT APPLICATION NUMBER: US/09/925,300  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1824  
LENGTH: 95  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (59)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (74)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (76)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (78)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (84)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (85)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-300-1824

Query Match 88.1%, Score 37; DB 9; Length 95;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRG 6  
Db 66 PRPPRG 71

RESULT 12  
US-10-424-599-191513  
Sequence 191513, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599



COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICANT: Patentin Release #1.0, Version #1.30  
FILING DATE: 10-Jun-2003  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/778,717  
FILING DATE: 12-DEC-1996  
APPLICATION NUMBER: JP 352225/1995  
FILING DATE: 28-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NORMAN F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 2084-031-0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-413-3000  
TELEFAX: 703-413-2220  
RELEVANT RESIDUES IN SEQ ID NO: 7: FROM 1 TO 209  
SEQUENCE DESCRIPTION: SEQ ID NO: 7:

Query Match 88.1%; Score 37; DB 12; Length 96;  
Best Local Similarity 85.7%; Pred. No. 2.6e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PRPPRG 7  
Db 54 PNPGR 60

RESULT 13  
US-10-425-114-52873  
; Sequence 52873, Application US/10425114  
; Publication No. US2004003488A1  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovalic, David K.  
; APPLICANT: Screen, Steven E.  
; APPLICANT: Tabaska, Jack E.  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/425,114  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 52873  
; LENGTH: 131  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3060-015-B9\_FLI.pap  
US-10-425-114-52873

Query Match 88.1%; Score 37; DB 12; Length 131;  
Best Local Similarity 85.7%; Pred. No. 3.4e+02;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PRPPRG 7  
Db 97 PRPGR 103

RESULT 14  
US-10-457-372-7  
; GENERAL INFORMATION:  
; APPLICANT: UENO, EIICHI  
; APPLICANT: NOBUYUKI, FUJII  
; APPLICANT: OKADA, MASAHISA  
; TITLE OF INVENTION: FUSED DNA SEQUENCE, FUSED PROTEIN  
; FILE REFERENCE: 38-21(5313)B  
; CURRENT APPLICATION NUMBER: US/10/457,372  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 21  
; SEQ ID NO 457372  
; LENGTH: 219  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3060-015-B9\_FLI.pap  
US-10-457-372-7

Query Match 88.1%; Score 37; DB 16; Length 219;  
Best Local Similarity 100.0%; Pred. No. 5.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRG 6  
Db 80 PRPGR 85

Search completed: April 6, 2004, 17:06:07  
Job time : 27.4766 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:56:34 ; Search time 10.271 Seconds  
(without alignments)  
35.185 Million cell updates/sec

Title: US-10-009-709-10

Perfect score: 42

Sequence: 1 PRPGR 7

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.\*

- 1: /cgn2\_6/ptodata/2/iaa/5A COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PTUS COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	42	100.0	207	4	US-09-252-991A-29505
2	42	100.0	418	4	US-09-252-991A-29452
3	42	100.0	559	4	US-09-252-991A-24480
4	38	90.5	377	4	US-09-252-991A-29332
5	37	88.1	40	4	US-09-314-268-106
6	37	88.1	155	4	US-09-252-991A-20099
7	37	88.1	182	4	US-09-252-991A-21772
8	37	88.1	208	4	US-09-252-991A-32430
9	37	88.1	209	4	US-08-778-717-7
10	37	88.1	226	3	US-09-195-868-30
11	37	88.1	355	3	US-08-483-533-41
12	37	88.1	355	4	US-09-283-471A-41
13	37	88.1	355	5	PCT-US91-06532-3
14	37	88.1	429	1	US-07-672-483-3
15	37	88.1	429	3	US-09-370-368-6
16	37	88.1	429	6	5182195-13
17	37	88.1	474	3	US-08-729-416C-1
18	37	88.1	474	3	US-08-729-416C-11
19	37	88.1	474	3	US-08-807-342B-2
20	37	88.1	474	4	US-09-433-353-1
21	37	88.1	474	4	US-08-433-353-11
22	37	88.1	594	3	US-08-729-416C-7
23	37	88.1	594	4	US-09-433-353-7
24	37	88.1	594	4	US-09-433-353-17
25	37	88.1	1229	3	US-09-195-868-28
26	37	88.1	1427	4	US-09-252-991A-20577
27	37	88.1	1427	4	Sequence 20577, A

28	37	88.1	1466	4	US-09-262-537-20	Sequence 20, Appl
29	37	88.1	1469	4	US-09-262-537-58	Sequence 58, Appl
30	37	88.1	1471	4	US-08-811-519-1	Sequence 1, Appl
31	36	85.7	199	4	US-09-252-991A-26006	Sequence 26006, A
32	36	85.7	208	4	US-09-252-991A-20806	Sequence 20806, A
33	36	85.7	245	4	US-09-252-991A-21078	Sequence 21078, A
34	36	85.7	297	4	US-09-252-991A-26587	Sequence 26587, A
35	36	85.7	418	4	US-09-252-991A-17796	Sequence 17796, A
36	36	85.7	447	1	US-07-937-609-29	Sequence 29, Appl
37	36	85.7	447	1	US-07-978-892A-6	Sequence 6, Appl
38	36	85.7	447	3	US-08-029-170-29	Sequence 29, Appl
39	36	85.7	448	1	US-08-570-157-3	Sequence 3, Appl
40	36	85.7	448	4	US-09-076-510-3	Sequence 3, Appl
41	36	85.7	448	4	US-09-004-349-3	Sequence 3, Appl
42	36	85.7	449	1	US-08-570-157-1	Sequence 1, Appl
43	36	85.7	449	4	US-09-076-510-1	Sequence 1, Appl
44	36	85.7	449	4	US-09-004-349-1	Sequence 1, Appl
45	36	85.7	451	1	US-08-570-157-2	Sequence 2, Appl

## ALIGNMENTS

### RESULT 1

US-09-252-991A-29505  
; Sequence 29505, Application US/09252991A  
; Patent No. 6851795

; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 29505

; LENGTH: 207

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-29505

Query Match 100.0%; Score 42; DB 4; Length 207;

Best Local Similarity 100.0%; Pred. No. 16;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPGR 7

Db 84 PRPGR 90

### RESULT 2

US-09-252-991A-29452

; Sequence 29452, Application US/09252991A

; Patent No. 6851795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 29452

; LENGTH: 418

; TYPE: PRT

ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29452

Query Match 100.0%; Score 42; DB 4; Length 418;  
Best Local Similarity 100.0%; Pred. No. 29;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
Db 161 PRPPGR 167

## RESULT 3

US-09-252-991A-24480  
Sequence 24480, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252.991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 24480

LENGTH: 559

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-24480

Query Match 100.0%; Score 42; DB 4; Length 559;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
Db 315 PRPPRGR 321

## RESULT 4

US-09-252-991A-29332

Sequence 29332, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252.991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 29332

LENGTH: 377

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-29332

Query Match 90.5%; Score 38; DB 4; Length 377;  
Best Local Similarity 85.7%; Pred. No. 1e-02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
Db 93 PRPPRGR 99

## RESULT 5

US-09-314-268-106

Sequence 106, Application US/09314268

Patent No. 6346377

GENERAL INFORMATION:

APPLICANT: Doorbar, John

TITLE OF INVENTION: IMPROVEMENTS IN OR RELATING TO SCREENING FOR PAPILLOMA

TITLE OF INVENTION: VIRUSES

FILE REFERENCE: 3789/80902

CURRENT APPLICATION NUMBER: US/09/314,268

CURRENT FILING DATE: 1999-03-19

EARLIER APPLICATION NUMBER: 09/314,268

EARLIER FILING DATE: 1999-05-18

NUMBER OF SEQ ID NOS: 179

SOFTWARE: Patentin ver. 2.1

SEQ ID NO 106

LENGTH: 40

TYPE: PRT

ORGANISM: Human papillomavirus type 30

US-09-314-268-106

Query Match 88.1%; Score 37; DB 4; Length 40;  
Best Local Similarity 85.7%; Pred. No. 20;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PRPPRGR 7  
Db 10 PRPPGR 16

## RESULT 6

US-09-252-991A-20099

Sequence 20099, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252.991A

CURRENT FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 20099

LENGTH: 155

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-20099

Query Match 88.1%; Score 37; DB 4; Length 155;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRGR 6  
Db 82 PRPPRGR 87

## RESULT 7

US-09-252-991A-21772

Sequence 21772, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252.991A

CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 21772  
 ; LENGTH: 182  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 JS-09-252-991A-21772

Query Match 88.1%; Score 37; DB 4; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 77;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRG 6  
 Db 108 PRPPRG 113

RESULT 8  
 US-09-252-991A-32430  
 ; Sequence 32430, Application US/09252991A  
 ; Patent No. 6851795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/094,190  
 ; PRIOR FILING DATE: 1998-07-27  
 ; NUMBER OF SEQ ID NOS: 33142  
 ; SEQ ID NO 32430  
 ; LENGTH: 208  
 ; TYPE: PRT  
 ; ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-32430

Query Match 88.1%; Score 37; DB 4; Length 208;  
 Best Local Similarity 85.7%; Pred. No. 87;  
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PRPPRG 7  
 Db 133 PRPPRG 139

RESULT 9  
 US-08-778-717-7  
 ; Sequence 7, Application US/08778717  
 ; Patent No. 6802689  
 ; GENERAL INFORMATION:  
 ; APPLICANT: UENO, EIICHI  
 ; APPLICANT: NOBUYUKI, FUJII  
 ; APPLICANT: OKADA, MASAHISA  
 ; TITLE OF INVENTION: FUSED DNA SEQUENCE, FUSED PROTEIN  
 ; TITLE OF INVENTION: EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR  
 ; TITLE OF INVENTION: EXPRESSING SAID DNA SEQUENCE  
 ; NUMBER OF SEQUENCES: 21  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 ; ADDRESSEE: P.C.  
 ; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
 ; CITY: ARLINGTON  
 ; STATE: VA  
 ; COUNTRY: USA  
 ; ZIP: 22202  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/778,717  
 ; FILING DATE: 12-DEC-1996  
 ; CLASSIFICATION: 536  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: JP 352225/1995  
 ; FILING DATE: 28-DEC-1995  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: OBLON, NORMAN F.  
 ; REGISTRATION NUMBER: 24,618  
 ; REFERENCE/DOCKET NUMBER: 2084-031-0  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 703-413-3000  
 ; TELEFAX: 703-413-2220  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 209 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: RECOMBINANT  
 ; PUBLICATION INFORMATION:  
 ; AUTHORS: NOBUYUKI FUJII ET AL,  
 ; TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED  
 ; TITLE: FROM SAID FUSED DNA SEQUENCE AND METHOD OF  
 ; TITLE: EXPRESSING SAID FUSED PROTEIN  
 ; RELEVANT RESIDUES IN SEQ ID NO: 7: FROM 1 TO 209  
 US-08-778-717-7

Query Match 88.1%; Score 37; DB 4; Length 209;  
 Best Local Similarity 100.0%; Pred. No. 87;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPPRG 6  
 Db 92 PRPPRG 97

RESULT 10  
 US-09-195-868-30  
 ; Sequence 30, Application US/09195868  
 ; Patent No. 6090621  
 ; GENERAL INFORMATION:  
 ; APPLICANT: KAVANAUGH MD, MICHAEL  
 ; APPLICANT: POT PH.D., DAVID  
 ; APPLICANT: WILLIAMS MDPH, LEWIS T.  
 ; TITLE OF INVENTION: SIGNALING INOSITOL POLYPHOSPHATE  
 ; NUMBER OF SEQUENCES: 33  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: CHIRON CORPORATION  
 ; STREET: 4560 HORTON STREET  
 ; CITY: EMERYVILLE  
 ; STATE: CA  
 ; COUNTRY: USA  
 ; ZIP: 94608  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/195,868  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; APPLICATION NUMBER: US/09/195,868  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:

NAME: FIRESTONE, LEIGH H.  
REGISTRATION NUMBER: 36,831  
REFERENCE/DOCKET NUMBER: 1182.004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 510-923-2707  
TELEFAX: 510-655-3542  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 226 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-195-868-30

Query Match 88.1%; Score 37; DB 3; Length 226;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPRPG 6  
Db 2 PRPRPG 7

RESULT 11  
US-08-483-533-41  
; Sequence 41, Application US/08483533  
; Patent No. 6172047  
; GENERAL INFORMATION:  
; APPLICANT: Roizman, Bernard  
; APPLICANT: Chou, Joany  
; TITLE OF INVENTION: Method for Treating Tumorigenic  
; DISEASES  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/483,533  
; FILING DATE: 07-MAR-95  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/419,853  
; FILING DATE: 11-APR-95  
; APPLICATION NUMBER: 07/861,233  
; FILING DATE: 31-MAR-92  
; APPLICATION NUMBER: 07/861,233  
; FILING DATE: 11-APR-95  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, James P.  
; REGISTRATION NUMBER: 28,491  
; REFERENCE/DOCKET NUMBER: 28097/32742  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 355 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-483-533-41

Query Match 88.1%; Score 37; DB 3; Length 355;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPRPG 6  
Db 173 PRPRPG 178

RESULT 12  
US-09-283-471A-41  
; Sequence 41, Application US/09283471A  
; Patent No. 6340673  
; GENERAL INFORMATION:  
; APPLICANT: Roizman, Bernard  
; APPLICANT: Chou, Joany  
; TITLE OF INVENTION: Method For Treating Tumorigenic Diseases  
; DISEASES  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 6300 Sears Tower, 233 South Wacker Drive  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: United States of America  
; ZIP: 60606-6402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/283,471A  
; FILING DATE: 04-APR-1999  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/861,233  
; FILING DATE: 31-MAR-1992  
; APPLICATION NUMBER: 08/419,853  
; FILING DATE: 11-APR-1995  
; APPLICATION NUMBER: 08/483,533  
; FILING DATE: 07-JUN-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Zeller, James P.  
; REGISTRATION NUMBER: 28,491  
; REFERENCE/DOCKET NUMBER: 27373/32742A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312/474-6300  
; TELEFAX: 312/474-0448  
; TELEX: 25-3856  
; INFORMATION FOR SEQ ID NO: 41:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 355 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-283-471A-41

Query Match 88.1%; Score 37; DB 4; Length 355;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPRPG 6  
Db 173 PRPRPG 178

RESULT 13  
PCT-US91-06532-3  
; Sequence 3, Application PC/TUS9106532  
; GENERAL INFORMATION:

APPLICANT: Roizman, Bernard  
TITLE OF INVENTION: Recombinant Herpes Simplex Viruses  
TITLE OF INVENTION: Vaccines and Methods  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
ADDRESSEE: Bicknell  
STREET: Two First National Plaza Suite 2100  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US91/06532  
FILING DATE: 19910910  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Gruber, Lewis S.  
REGISTRATION NUMBER: 20,060  
REFERENCE/DOCKET NUMBER: 27373/8235  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312/346-5750  
TELEFAX: 312/984-9740  
TELEX: 25-3856  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 355 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US91-06532-3

Query Match 88.1%; Score 37; DB 5; Length 355;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPRPG 6  
DB 173 PRPRPG 178

RESULT 14  
US-07-672-483-3  
Sequence 3, Application US/07672483  
Patent No. 5359029  
GENERAL INFORMATION:  
APPLICANT: LACROIX, Martial  
APPLICANT: ZREIN, Maan  
TITLE OF INVENTION: PEPTIDES AND ANALOGUES AND MIXTURES  
TITLE OF INVENTION: THEREOF FOR DETECTING ANTIBODIES TO HTLV-I AND HTLV-II  
TITLE OF INVENTION: VIRUSES  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FISH & NEAVE  
STREET: 875 Third Avenue  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10022-6250  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/672,483  
FILING DATE: 19910302  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/554,258  
FILING DATE: 18-JUL-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: HALEY Jr, James F  
REGISTRATION NUMBER: 27,794  
REFERENCE/DOCKET NUMBER: IAF8 CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212)715-0600  
TELEFAX: (212)715-0674  
TELEX: 14-8367  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 429 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-672-483-3

Query Match 88.1%; Score 37; DB 1; Length 429;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPRPG 6  
DB 13 PRPRPG 18

RESULT 15  
US-09-370-368-6  
Sequence 6, Application US/09370368  
Patent No. 6258932  
GENERAL INFORMATION:  
APPLICANT: Anders Vahne  
TITLE OF INVENTION: PEPTIDES THAT BLOCK VIRAL INFECTIVITY  
FILE REFERENCE: TRIPEP.003A  
CURRENT APPLICATION NUMBER: US/09/370,368  
CURRENT FILING DATE: 1999-08-09  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6  
LENGTH: 429  
TYPE: PRT  
ORGANISM: Human T-Cell Lymphotropic Virus Type 1  
US-09-370-368-6

Query Match 88.1%; Score 37; DB 3; Length 429;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PRPRPG 6  
DB 13 PRPRPG 18

Search completed: April 6, 2004, 16:19:41  
Job time : 10.271 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:52:34 ; Search time 6.72897 Seconds  
(without alignments)  
85.771 Million cell updates/sec

Title: US-10-009-709-11

Perfect score: 35

Sequence: 1 RPRGR 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78.\*

1: Pirl.\*

2: Pirl2.\*

3: Pirl3.\*

4: Pirl4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	166	1	PIHUSC
2	35	100.0	166	2	salivary proline-r
3	35	100.0	171	2	proline-rich phosph
4	35	100.0	433	2	hypothetical prote
5	35	100.0	561	2	hypothetical prote
6	35	100.0	784	2	exoribonuclease li
7	35	100.0	784	2	exoribonuclease RN
8	32	91.4	167	2	hypothetical prote
9	32	91.4	200	2	hypothetical prote
10	32	91.4	363	2	histidine-rich kno
11	32	91.4	541	2	hypothetical prote
12	32	91.4	614	1	probable adenyl-l
13	32	91.4	669	2	hepatoma-derived g
14	32	91.4	840	2	hypothetical prote
15	32	91.4	1429	2	nitric-oxide synth
16	32	91.4	1433	2	bone morphogenetic
17	31	88.6	36	2	probable deacetyla
18	31	88.6	260	2	RNA-binding protei
19	31	88.6	334	2	RNA-binding protei
20	31	88.6	342	2	nitrilotriacetate
21	31	88.6	450	2	prunin 1 precursor
22	31	88.6	551	2	procollagen C-endo
23	31	88.6	730	1	BMH1
24	31	88.6	787	2	D75325
25	31	88.6	1153	2	T00249
26	30	85.7	55	2	DNA-binding protei
27	30	85.7	55	2	hypothetical prote
28	30	85.7	79	2	homeotic protein C
29	30	85.7	100	2	hypothetical prote

30 85.7 108 1 B3AG55  
31 85.7 123 2 D72579  
32 85.7 128 2 C55224  
33 85.7 129 2 H72598  
34 85.7 132 2 B85092  
35 85.7 137 2 A70913  
36 85.7 137 2 A88637  
37 85.7 143 2 E72699  
38 85.7 150 2 D72670  
39 85.7 151 2 H82546  
40 85.7 168 2 T46341  
41 85.7 170 2 T15991  
42 85.7 178 2 A75036  
43 85.7 179 2 T35964  
44 85.7 192 2 A72646  
45 85.7 198 2 S72709

virB3 protein - Ag  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
protein W09G12.9 [protein W09G12.9]  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
hypothetical prote  
Lepb1170\_C3\_229 pr

#### ALIGNMENTS

##### RESULT 1

PIHUSC

salivary proline-rich phosphoprotein precursor PRH2 [validated] - human

N/Alternate names: salivary acidic proline-rich protein PRH2

C/Contains: Peptide P-C (basic proline-rich peptide 1B-8b); proline-rich phosphoprotein

C/Species: Homo sapiens (man)

C/Date: 31-Mar-1981 #sequence, revision 12-Apr-1996 #text change 08-Dec-2000

C/Accession: A25372; A19803; B57868; A92277; A92254; A94425; A91954; S02564; S02563; JPO

R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.

J. Biol. Chem. 260, 11123-11130, 1985

A/Title: Differential RNA splicing and post-translational cleavages in the human salivar

A/Reference number: A92492; MUID:85289325; PMID:2993301

A/Accession: A25372

A/Molecule type: mRNA

A/Residues: 1-166 <MAE>

A/Cross-references: GB:K03202; NID:G190481; PIDN:AAA60183.1; PID:G190482

R/Schlesinger, D.H.; Hay, D.I.

Int. J. Pept. Protein Res. 17, 34-41, 1981

A/Title: Primary structure of the active tryptic fragments of human and monkey salivary

A/Reference number: A91757; MUID:81191179; PMID:7228490

A/Accession: A19803

A/Molecule type: protein

A/Residues: 17-46 <SCH>

R/Kim, H.S.; Maeda, N.

J. Biol. Chem. 261, 6712-6718, 1986

A/Title: Structures of two HaeIII-type genes in the human salivary proline-rich protein

A/Reference number: A57868; MUID:86196106; PMID:3009472

A/Accession: B57868

A/Molecule type: DNA

A/Residues: 1-166 <KIM>

A/Cross-references: GB:M13058; NID:G190513; PIDN:AAA98808.1; PID:G190514

R/Wong, R.S.C.; Bennick, A.

J. Biol. Chem. 255, 5943-5948, 1980

A/Title: The primary structure of a salivary calcium-binding proline-rich phosphoprotein

A/Reference number: A92277; MUID:80204368; PMID:7380845

A/Contents: protein C

A/Accession: A92277

A/Molecule type: protein

A/Residues: 17-19, 'N', 21-166 <WON>

A/Note: the amino-terminal 46 residues are involved with inhibiting hydroxyapatite forma

R/Wong, R.S.C.; Hofmann, T.; Bennick, A.

J. Biol. Chem. 254, 4800-4808, 1979

A/Title: The complete primary structure of a proline-rich phosphoprotein from human sali

A/Reference number: A92254; MUID:79173237; PMID:438215

A/Contents: protein A

A/Accession: A92254

A/Molecule type: protein

A/Residues: 17-19, 'N', 21-122 <WO2>

R/Schlesinger, D.H.; Hay, D.I.

in Peptides: Structure and Biological Function (Proc. Sixth Amer. Peptide Symp.), Gross,

A/Title: Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent i

A/Reference number: A94425

A;Accession: A94425  
 A;Molecule type: protein  
 A;Residues: 17-122 <SC2>  
 A;Note: the authors call this protein PRP-4  
 R;Iseumura, S.; Saitoh, E.; Sanada, K.  
 J. Biochem. 87, 1071-1077, 1980  
 A;Title: The amino acid sequence of a salivary proline-rich peptide, P-C, and its relation to the human salivary proline-rich protein PRP-4  
 A;Reference number: A91954; MUID:80227634; PMID:7390979  
 A;Contents: peptide P-C  
 A;Accession: A91954  
 A;Molecule type: protein  
 A;Residues: 123-166 <ISE>  
 R;Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck  
 Biochem. J. 255, 15-21, 1988  
 A;Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-1, PRP-2, PRP-3, PRP-4, PRP-5, PRP-6)  
 A;Reference number: S02562; MUID:89061650; PMID:3196309  
 A;Accession: S02564  
 A;Molecule type: protein  
 A;Residues: 17-166 <HAY>  
 A;Accession: S02563  
 A;Molecule type: protein  
 A;Residues: 47-71 <HA2>  
 R;Schlesinger, D.H.; Hay, D.I.  
 Int. J. Pept. Protein Res. 27, 373-379, 1986  
 A;Title: Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of the human salivary proline-rich protein PRP-2  
 A;Reference number: JP0106; MUID:86222916; PMID:3710693  
 A;Accession: JP0106  
 A;Molecule type: protein  
 A;Residues: 17-161, Q', 163-166 <SC3>  
 A;Experimental source: parotid gland  
 R;Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A;Title: Basic proline-rich proteins from human parotid saliva: relationships of the cov  
 A;Reference number: A38355; MUID:91190884; PMID:1849422  
 A;Accession: G38355  
 A;Status: preliminary  
 A;Molecule type: protein  
 A;Residues: 123-166 <KAU>  
 R;Robinson, R.; Kaufman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A;Title: Primary structure and possible origin of the non-glycosylated basic proline-rich  
 A;Reference number: S06153; MUID:90088384; PMID:2688632  
 A;Accession: S06153  
 A;Molecule type: protein  
 A;Residues: 123-166 <ROB>  
 C;Comment: The proposed biological functions are a highly potent inhibitor of crystal gr  
 C;Genetics:  
 A;Gene: GDB:PRH2  
 A;Cross-references: GDB:119516; OMIM:168790  
 A;Map position: 12p13.2-12p13.2  
 A;Intons: 22/1; 34/1  
 C;Superfamily: proline-rich protein  
 C;Keywords: calcium binding; phosphoprotein; pyroglutamic acid; saliva  
 F;1-16/Domain: signal sequence #status predicted <SIG>  
 F;17-166/Product: protein C #status experimental <PRC>  
 F;17-122/Product: protein A #status experimental <PRA>  
 F;17-46/Region: apatitic mineral binding  
 F;17-71/Product: PRP-3 #status experimental <PRP3>  
 F;123-166/Product: peptide P-C #status experimental <PPC>  
 F;17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experimen  
 F;24,38/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 100.0%; Score 35; DB 1; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
 DB 119 RPPRGR 124

RESULT 2  
 B25372

salivary proline-rich phosphoprotein precursor PRH1 (allele PIF) - human  
 C;Species: Homo sapiens (man)  
 C;Date: 29-Aug-1987 #sequence revision 29-Aug-1987 #text change 20-Aug-1999  
 A;Accession: B25372; A57868; S02562; G38355; S06153; B27307  
 R;Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
 J. Biol. Chem. 260, 11123-11130, 1985  
 A;Title: Differential RNA splicing and post-translational cleavages in the human salivary  
 A;Reference number: A92492; MUID:85289325; PMID:2993301  
 A;Accession: B25372  
 A;Molecule type: mRNA  
 A;Residues: 1-166 <MAE>  
 A;Cross-references: GB:K03203; NID:G190483; PIDN:AAA60184.1; PID:G190484  
 R;Kim, H.S.; Maeda, N.  
 J. Biol. Chem. 261, 6712-6718, 1986  
 A;Title: Structures of two HaeIII-type genes in the human salivary proline-rich protein  
 A;Reference number: A57868; MUID:86196106; PMID:3009472  
 A;Accession: A57868  
 A;Molecule type: DNA  
 A;Residues: 1-166 <KIM>  
 A;Cross-references: GB:M13057; NID:G190511; PIDN:AAA98807.1; PID:G190512  
 R;Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck  
 Biochem. J. 255, 15-21, 1988  
 A;Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-1, PRP-2, PRP-3, PRP-4, PRP-5, PRP-6)  
 A;Reference number: S02562; MUID:89061650; PMID:3196309  
 A;Accession: S02562  
 A;Molecule type: protein  
 A;Residues: 47-71 <HAY>  
 R;Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A;Title: Basic proline-rich proteins from human parotid saliva: relationships of the cov  
 A;Reference number: A38355; MUID:91190884; PMID:1849422  
 A;Accession: G38355  
 A;Molecule type: protein  
 A;Residues: 123-166 <KAU>  
 R;Robinson, R.; Kaufman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A;Title: Primary structure and possible origin of the non-glycosylated basic proline-rich  
 A;Reference number: S06153; MUID:90088384; PMID:2688632  
 A;Accession: S06153  
 A;Molecule type: protein  
 A;Residues: 123-166 <ROB>  
 R;Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A;Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich pro  
 A;Reference number: A27307; MUID:88074309; PMID:3687941  
 A;Contents: allele Pa  
 A;Accession: B27307  
 A;Status: nucleic acid sequence not shown  
 A;Molecule type: DNA  
 A;Residues: 17-41, 'L', 43-118, 'C', 120-166 <AZB>  
 A;Cross-references: EMBL:K03203  
 C;Genetics:  
 A;Gene: GDB:PRH1  
 A;Cross-references: GDB:119515; OMIM:168730  
 A;Map position: 12p13.2-12p13.2  
 C;Superfamily: proline-rich protein  
 C;Keywords: phosphoprotein; saliva; tandem repeat  
 Query Match 100.0%; Score 35; DB 2; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
 DB 119 RPPRGR 124

RESULT 3  
 A27307  
 N;Alternate names: salivary acidic proline-rich protein  
 C;Species: Homo sapiens (man)



1;Date: 30-Jun-1998 #sequence\_revision 30-Jun-1998 #text\_change 29-Aug-1997  
 2;Accession: A27307  
 3;Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 4;Ann. J. Hum. Genet. 41, 1035-1047, 1987  
 5;A;Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich pro  
 6;A;Reference number: A27307; PMID:88074309; PMID:3687941  
 7;Accession: A27307  
 8;A;Status: nucleic acid sequence not shown  
 9;A;Molecule type: DNA  
 10;Residues: 1-171 <AZE>  
 11;Cross-references: EMBL:K03203  
 12;Genetics:  
 13;A;Gene: GDB:PRH1  
 14;A;Cross-references: GDB:119515; OMIM:168730  
 15;A;Map position: 12p13.2-12p13.2  
 16;A;Superfamily: proline-rich protein  
 17;Keywords: phosphoprotein  
  
 Query Match 100.0%; Score 35; DB 2; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 2y 1 RPPRGR 6  
 |||||  
 Db 124 RPPRGR 129  
  
 RESULT 4  
 D84335  
 Hypothetical protein Vng1842h [imported] - Halobacterium sp. NRC-1  
 C;Species: Halobacterium sp. NRC-1  
 C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
 C;Accession: D84335  
 R;Ng, W.V.; Kennedy, S.P.; Manaitas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.  
 ; Lethaus, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabil  
 Jung, K.H.; Alam, M.; Freitas, T.  
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
 A;Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li  
 A;Title: Genome sequence of Halobacterium species NRC-1.  
 A;Reference number: A84160; PMID:20504483; PMID:11016950  
 A;Accession: D84335  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-433 <STO>  
 A;Cross-references: GB:AE004437; NID:gl0581288; PIDN:AG20048.1; GSPDB:GN00138  
 C;Genetics:  
 A;Gene: Vng1842h  
  
 Query Match 100.0%; Score 35; DB 2; Length 433;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
 2y 1 RPPRGR 6  
 |||||  
 Db 71 RPPRGR 76  
  
 RESULT 5  
 E70610  
 Hypothetical protein Rv1215c - Mycobacterium tuberculosis (strain H37RV)  
 C;Species: Mycobacterium tuberculosis  
 C;Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999  
 C;Accession: E70610  
 R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.  
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.  
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A;Reference number: A70500; PMID:9825987; PMID:9634230  
 A;Accession: E70610  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA

A;Residues: 1-561 <COL>  
 A;Cross-references: GB:293777; GB:AL123456; NID:g3261726; PIDN:CA807817.1; PID:e311160;  
 A;Experimental source: strain H37RV  
 C;Genetics:  
 A;Gene: Rv1215c

Query Match 100.0%; Score 35; DB 2; Length 561;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
 |||||  
 Db 536 RPPRGR 541

## RESULT 6

AG2736  
 exoribonuclease [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C;Species: Agrobacterium tumefaciens  
 C;Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C;Accession: AG2736  
 R;Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A;Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 ster, E.W.  
 A;Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A;Reference number: AB2577; PMID:21608550; PMID:11743193  
 A;Accession: AG2736  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-784 <KUR>  
 A;Cross-references: GB:AE008688; PIDN:AAL42309.1; PID:gl7739711; GSPDB:GN00186  
 A;Experimental source: strain C58 (Dupont)  
 C;Genetics:  
 A;Gene: rnr  
 A;Map position: circular chromosome  
 C;Superfamily: virulence-associated protein vacB homolog

Query Match 100.0%; Score 35; DB 2; Length 784;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
 |||||  
 Db 778 RPPRGR 783

## RESULT 7

E97517  
 exoribonuclease RNase R (PA4937) [imported] - Agrobacterium tumefaciens (strain C58, Cer  
 C;Species: Agrobacterium tumefaciens  
 C;Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 18-Nov-2002  
 C;Accession: E97517  
 R;Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
 A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.;  
 Science 294, 2323-2328, 2001  
 A;Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum  
 A;Reference number: A97359; PMID:21608551; PMID:11743194  
 A;Accession: E97517  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-784 <KUR>  
 A;Cross-references: GB:AE007869; PIDN:AAK87094.1; PID:gl5156354; GSPDB:GN00169  
 C;Genetics:  
 A;Gene: AGR\_C2395  
 A;Map position: circular chromosome  
 C;Superfamily: virulence-associated protein vacB homolog

Query Match 100.0%; Score 35; DB 2; Length 784;  
 Best Local Similarity 100.0%; Pred. No. 69;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
|||||  
Db 778 RPPRGR 783

RESULT 8  
A:Title: hypothetical protein Atu0440 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
A:Accession: AB2630  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
C:Accession: AB2630  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L.  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell  
; Karp, P.; Romero, E.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; MUID:21608550; PMID:11743193  
A:Accession: AB2630  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-167 <K>  
A:Cross-references: GB:AE008688; PIDN:AA141459.1; PID:G17738783; GSPDB:GN00186  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: Atu0440  
A:Map position: circular chromosome

Query Match 91.4%; Score 32; DB 2; Length 167;  
Best Local Similarity 83.3%; Pred. No. 61;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
|||||  
Db 68 KPPRGR 73

RESULT 9  
S55609  
hypothetical protein 13 - equine herpesvirus 2  
C:Species: equine herpesvirus 2  
C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 08-Oct-1999  
C:Accession: S55609  
R:Telord, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.  
J. Mol. Biol. 249, 520-528, 1995  
A:Title: The DNA sequence of equine herpesvirus 2.  
A:Reference number: S55594; MUID:95302501; PMID:7783207  
A:Accession: S55609  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-200 <TEL>  
A:Cross-references: GB:U20824; NID:G695172; PIDN:AAC13802.1; PID:G695187  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

Query Match 91.4%; Score 32; DB 2; Length 200;  
Best Local Similarity 83.3%; Pred. No. 72;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
|||||  
Db 14 KPPRGR 19

RESULT 10  
C44971  
histidine-rich knob protein homolog XPRC - Plasmodium cynomolgi (fragment)  
C:Species: Plasmodium cynomolgi  
C:Date: 14-May-1993 #sequence\_revision 14-May-1993 #text\_change 03-Nov-2000  
C:Accession: C44971  
R:Killejian, A.; Yang, Y.F.; Cochrane, A.H.; Rashid, M.A.  
Mol. Biochem. Parasitol. 38, 291-293, 1990

A:Title: Homologous sequences in Plasmodium cynomolgi and the gene of the histidine-rich  
A:Reference number: A44971; MUID:90220761; PMID:2183051  
A:Accession: C44971  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-363 <KIL>  
A:Cross-references: GB:M28063; NID:G160330; PID:G951371  
C:Superfamily: knob-associated histidine-rich protein

Query Match 91.4%; Score 32; DB 2; Length 363;  
Best Local Similarity 83.3%; Pred. No. 1.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
|||||  
Db 272 KPPRGR 277

RESULT 11  
F96776  
hypothetical protein F25A4.29 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Apr-2001  
C:Accession: F96776  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.,  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luroe, J.S.; Maiti, R.; Marziali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakanc, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: F96776  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-541 <STO>  
A:Cross-references: GB:AE005173; NID:G5882721; PIDN:AA055274.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F25A4.29  
A:Map position: 1  
C:Superfamily: calcium-dependent protein kinase; calmodulin repeat homology; protein kin  
C:Keywords: EF hand

Query Match 91.4%; Score 32; DB 2; Length 541;  
Best Local Similarity 83.3%; Pred. No. 1.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
|||||  
Db 22 KPPRGR 27

RESULT 12  
B70772  
probable adenylyl-sulfate kinase (EC 2.7.1.25) - Mycobacterium tuberculosis (strain H37F  
N:Alternate names: ATP sulfurylase large chain  
N:Contains: adenylylsulfate kinase (EC 2.7.1.25); sulfate adenylyltransferase (EC 2.7.7  
C:Species: Mycobacterium tuberculosis  
C:Date: 21-May-1999 #sequence\_revision 21-May-1999 #text\_change 03-Jun-2002  
C:Accession: B70772  
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S  
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998  
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
A:Reference number: A70500; MUID:98295987; PMID:9634230  
A:Accession: B70772  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:51:34 ; Search time 21.7009 Seconds  
(without alignments)  
87.236 Million cell updates/sec

Title: US-10-009-709-11

Perfect score: 35

Sequence: 1 RPRGR 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	186	16 Q82GY4	Q82GY4 streptomyc
2	35	100.0	246	16 Q9L0J9	Q9L0J9 streptomyc
3	35	100.0	255	10 Q94B27	Q94B27 arabidopsi
4	35	100.0	266	16 Q86639	Q86639 streptomyc
5	35	100.0	267	16 Q92K61	Q92K61 streptomyc
6	35	100.0	332	2 Q54729	Q54729 synchococ
7	35	100.0	344	16 Q8P1T0	Q8P1T0 xanthomon
8	35	100.0	378	4 Q96LN6	Q96LN6 homo sapien
9	35	100.0	420	17 Q8TXV0	Q8TXV0 methanopyru
10	35	100.0	433	17 Q9HP22	Q9HP22 halobacteri
11	35	100.0	535	10 Q84SD0	Q84SD0 oryza sativ
12	35	100.0	561	16 Q05316	Q05316 mycobacteri
13	35	100.0	561	16 Q70UD5	Q70UD5 mycobacteri
14	35	100.0	591	10 Q8S103	Q8S103 oryza sativ
15	35	100.0	658	10 Q7YCD7	Q7YCD7 oryza sativ
16	35	100.0	696	4 Q8NBE6	Q8NBE6 homo sapien

17	35	100.0	750	12 Q9YQ39	Q9YQ39 pseudorabi
18	35	100.0	784	16 Q8UFU4	Q8UFU4 agrobacteri
19	35	100.0	789	16 Q92QMO	Q92QMO rhizobium m
20	35	100.0	933	16 Q7UXE6	Q7UXE6 rhodospirell
21	35	100.0	975	5 Q8MZE9	Q8MZE9 drosophila
22	35	100.0	975	5 Q8ML73	Q8ML73 drosophila
23	35	100.0	1384	2 Q9AJPS	Q9AJPS myxococcus
24	35	100.0	1694	10 Q9FFH1	Q9FFH1 arabidopsi
25	35	100.0	5017	2 Q8GEX6	Q8GEX6 polyangium
26	32	91.4	167	16 Q8UI60	Q8UI60 agrobacteri
27	32	91.4	200	12 Q66619	Q66619 equine herp
28	32	91.4	214	4 Q8BYD4	Q8BYD4 homo sapien
29	32	91.4	245	4 Q86Q18	Q86Q18 homo sapien
30	32	91.4	339	4 Q86Z13	Q86Z13 homo sapien
31	32	91.4	363	5 Q25680	Q25680 plasmodium
32	32	91.4	397	11 Q9EQJ1	Q9EQJ1 rattus norv
33	32	91.4	460	4 Q7Z722	Q7Z722 homo sapien
34	32	91.4	478	16 Q92RY1	Q92RY1 rhizobium m
35	32	91.4	486	4 Q96DX9	Q96DX9 homo sapien
36	32	91.4	486	11 Q8R1C9	Q8R1C9 mus musculu
37	32	91.4	504	16 Q92LD7	Q92LD7 rhizobium m
38	32	91.4	517	11 Q8C3E0	Q8C3E0 mus musculu
39	32	91.4	541	10 Q9S8F8	Q9S8F8 arabidopsi
40	32	91.4	569	4 Q96G15	Q96G15 homo sapien
41	32	91.4	614	16 Q7U089	Q7U089 mycobacteri
42	32	91.4	669	11 Q35540	Q35540 mus musculu
43	32	91.4	669	11 Q925G1	Q925G1 rattus norv
44	32	91.4	670	4 Q9BW08	Q9BW08 homo sapien
45	32	91.4	671	4 Q7Z4V5	Q7Z4V5 homo sapien

## ALIGNMENTS

### RESULT 1

Q82GY4 PRELIMINARY; PRT; 186 AA.

AC Q82GY4;  
DT 01-JUN-2003 (TRENBLrel. 24, Created)  
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
DE Hypothetical protein.  
GN SAV3762.  
OS Streptomyces avermitilis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_taxid=33903;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=21477403; PubMed=11572948;  
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,  
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,  
RA Kikuchi H., Shiba I., Sakaki Y., Hattori M.;  
RT "Genome sequence of an industrial microorganism Streptomyces  
RT avermitilis: deducing the ability of producing secondary  
RT metabolites";  
RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=2260306; PubMed=12692562;  
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
RA Sakaki Y., Hattori M., Omura S.;  
RT "Complete genome sequence and comparative analysis of the industrial  
RT microorganism Streptomyces avermitilis";  
RL Nat. Biotechnol. 21:526-531(2003).  
DR EMBL; AP005036; BAC71474.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 186 AA; 20723 MW; 89128D7AC3CDB0C1 CRC64;

Query Match 100.0%; Score 35; DB 16; Length 186;  
Best Local Similarity 100.0%; Pred. No. 37;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPPRGR 6  
 DB 12 RPPRGR 17

## RESULT 2

Q9L0J9 ID O9L0J9 PRELIMINARY; PRT; 246 AA.  
 AC O9L0J9;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DE Putative ABC-transporter ATP-binding protein.  
 GN SCO4666 OR SCD40A.12C OR SCD40A.12C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Seeger K.J., Harris D.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,  
 RA Kinashi H., Hopwood D.A.;  
 RT "A set of ordered cosmids and a detailed genetic and physical map for  
 RT the 8 Mb Streptomyces coelicolor A3(2) chromosome."  
 RL Mol. Microbiol. 21:77-96(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2);  
 RC MEDLINE=97000351; PubMed=8843436;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,  
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Goble L., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2)."  
 RL Nature 417:141-147(2002).  
 CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.  
 DR EMBL; AL939121; CAB81857.1; .  
 DR GO; GO:0016020; C:membrane; IEA.  
 DR GO; GO:005524; F:ATP binding; IEA.  
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.  
 DR GO; GO:0006167; F:nucleotide binding; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR003593; AAA ATPase.  
 DR InterPro; IPR003439; ABC transporter.  
 DR Pfam; PF00005; ABC trans.1.  
 DR ProDom; PD000006; ABC transporter; 1.  
 DR SMART; SM00382; AAA; 1.  
 DR PROSITE; PS00211; ABC TRANSPORTER 1; 1.  
 DR PROSITE; PS00893; ABC TRANSPORTER 2; 1.  
 KW ATP-binding; Transport; Complete proteome.  
 SQ SEQUENCE 246 AA; 26164 MW; 2EBFF9753BE0AB62 CRC64;

Query Match 100.0%; Score 35; DB 16; Length 246;

Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPPRGR 6  
 DB 119 RPPRGR 124

## RESULT 3

Q94B27 ID Q94B27 PRELIMINARY; PRT; 255 AA.  
 AC Q94B27;  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DE Unknown protein (Hypothetical protein).  
 GN MLN1.10.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY042893; AAK68833.1; .  
 DR EMBL; AY072465; AAL66880.1; .  
 KW Hypothetical protein.  
 SQ SEQUENCE 255 AA; 27970 MW; C12D1E00235D4AF2 CRC64;

Query Match 100.0%; Score 35; DB 10; Length 255;  
 Best Local Similarity 100.0%; Pred. No. 50;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
 DB 32 RPPRGR 37

## RESULT 4

OS8639 ID OS8639 PRELIMINARY; PRT; 266 AA.  
 AC OS8639;  
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
 DE Hypothetical protein SC05719.  
 GN SC05719 OR SC3C3.05C.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Streptomycineae; Streptomycetaceae; Streptomyces.  
 OX NCBI\_TaxID=1902;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=A3(2) / M145;  
 RX MEDLINE=21996410; PubMed=12000953;  
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,

RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,  
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,  
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
 RA Warren T., Wietzorrek A., Woodward J., Barrall B.G., Parkhill J.,  
 RA Hopwood D.A.;  
 RT "Complete genome sequence of the model actinomycete Streptomyces  
 RT coelicolor A3(2).";  
 RL Nature 417:141-147(2002).  
 DR EMBL; AL939124; CAA20254.1; -;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 266 AA; 27842 MW; F2EP94BP1490E3DB CRC64;

Query Match 100.0%; Score 35; DB 16; Length 266;  
 Best Local Similarity 100.0%; Pred. NO. 52;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRCR 6

DB 166 RPPRCR 171

# RESULT 5

Q82K61 PRELIMINARY; PRT; 267 AA.

AC Q82K61;

DT 01-JUN-2003 (TREMBlrel. 24, Created)

DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)

DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)

DE Hypothetical protein.

GN SAV2543.

OS Streptomyces avermitilis.

OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;

OC Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI\_TaxID=33903;

RN [1]\_TaxID=33903;

RP SEQUENCE FROM N.A.

RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;

RX MEDLINE=21477403; PubMed=11572949;

RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,

RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,

RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,

RT "Genome sequence of an industrial microorganism Streptomyces

RT avermitilis: deducing the ability of producing secondary

RT metabolites.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;

RX MEDLINE=22608306; PubMed=12692562;

RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,

RA Sakaki Y., Hattori M., Omura S.,

RT "Complete genome sequence and comparative analysis of the industrial

RT microorganism Streptomyces avermitilis";

RL Nat. Biotechnol. 21:526-531(2003).

DR EMBL; AF005031; BAC70254.1; -;

DR InterPro; IPR008996; Cytok Ili like.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 267 AA; 27932 MW; 7B10A1B389C67F8 CRC64;

Query Match

Best Local Similarity 100.0%; Score 35; DB 16; Length 267;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRCR 6

DB 167 RPPRCR 172

# RESULT 6

Q54729

ID Q54729 PRELIMINARY; PRT; 332 AA.

AC Q54729;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)  
 DE ORF11692 (Fragment).  
 OS Synechococcus sp. (strain PCC 7942) (Anacystis nidulans R2).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
 OX NCBI\_TaxID=1140;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=PCC 7942;  
 RA Tsinoerenas N.F., Golden S.S.;  
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U44761; AAA86647.1; -;  
 DR InterPro; IPR001064; Crystallin.  
 DR PROSITE; PS00225; CRYSTALLIN\_BETAGAMMA; 1.  
 FT NON\_TER 1  
 SQ SEQUENCE 332 AA; 38574 MW; CA17B5F4B0F2ED09 CRC64;

Query Match 100.0%; Score 35; DB 2; Length 332;  
 Best Local Similarity 100.0%; Pred. NO. 64;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRCR 6

DB 283 RPPRCR 288

# RESULT 7

Q8PIT0

ID Q8PIT0 PRELIMINARY; PRT; 344 AA.

AC Q8PIT0;

DT 01-OCT-2002 (TREMBlrel. 22, Created)

DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)

DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)

DE Hypothetical protein XAC2815.

GN XAC2815.

OS Xanthomonas axonopodis (pv. citri).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;

OC Xanthomonadaceae; Xanthomonas.

OX NCBI\_TaxID=92829;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=306 / ATCC 13902 / XV 101;

RX MEDLINE=22022145; PubMed=1204217;

RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furlan L.R.,

RA Quaggio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,

RA Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,

RA Camarotte G., Cannavan F., Cardoso J., Chamargo F., Clapina L.P.,

RA Cicarelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorry H.,

RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,

RA Fornighieri E.F., Franco M.C., Greggio C.C., Gruber A.,

RA Katsuyama A.M., Kishii L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,

RA Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,

RA Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,

RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,

RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,

RA Spindola L.A.F., Takita M.A., Tamura R.B., Teixeira E.C., Tezza R.I.D.,

RA Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,

RA Setubal J.C., Kikajima J.P.;

RT "Comparison of the genomes of two Xanthomonas pathogens with differing

RT host specificities";

RL Nature 417:459-463(2002).

DR EMBL; AF011923; AAM37660.1; -;

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 344 AA; 38396 MW; 70020FA3B0C80B4C CRC64;

Query Match 100.0%; Score 35; DB 16; Length 344;

Best Local Similarity 100.0%; Pred. NO. 66;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRCR 6

DB 1 RPPRCR 6

Db 31 RPPRGR 36

RESULT 8

Q96LNG PRELIMINARY; PRT; 378 AA.  
AC Q96LNG;  
DT 01-DEC-2001 (TREMELrel. 19, Created)  
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)  
DE Hypothetical protein FLJ25348.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
CX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis.  
RA Isibashi T., Kanehori K., Yosida M., Watanabe S., Ishida S., Ono Y.,  
RA Hotuta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Chiba Y.,  
RA Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,  
RA Suzuki Y., Hata H., Nakagawa K., Mizuno S., Morinaga M., Kawamura M.,  
RA Sugiyama T., Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A.,  
RA Kawakami B., Nagai K., Isogai T., Sugano S.;  
RA "NEDO human cDNA sequencing project."  
RT Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL; AK058077; BAB7152.1; -.  
RW Hypothetical protein.  
KW SEQUENCE 378 AA; 41324 MW; D9ACAA5CACFDEB88C CRC64;  
SQ

Query Match 100.0%; Score 35; DB 4; Length 378;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6

Db 250 RPPRGR 255

RESULT 9

Q8TXV0 PRELIMINARY; PRT; 420 AA.  
AC Q8TXV0;  
DT 01-JUN-2002 (TREMELrel. 21, Created)  
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)  
DT 01-OCT-2002 (TREMELrel. 22, Last annotation update)  
DE Protein implicated in ribosomal biogenesis, Nop56p homolog.  
GN SIK1 OR MK0559.  
OS Methanopyrus kandleri.  
OC Archaea; Euryarchaeota; Methanopyri; Methanopyrales; Methanopyraceae;  
OC Methanopyrus.  
CX NCBI\_TaxID=2320;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=AV19 / DSM 6324 / JCM 9639;  
RX MEDLINE=21927647; PubMed=11930014;  
RA Slesarev A.I., Mezheva K.V., Makarova K.S., Polushin N.N.,  
RA Shcherbinina O.V., Shakhova V.V., Belova G.I., Aravind L.,  
RA Natale D.A., Rogozin I.B., Tatusov R.I., Wolf Y.I., Stetter K.O.,  
RA Malykh A.G., Koonin E.V., Kozlovskiy S.A.;  
RT "The complete genome of hyperthermophile Methanopyrus kandleri AV19  
RT and monophyly of archaeal methanogens."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:4644-4649 (2002).  
DR EMBL; AE010350; AA001774.1; -.  
DR InterPro; IPR002687; Nop.  
DR Pfam; PF01798; Nop; 1.  
DR ProDom; PD004104; Nop; 1.  
KW Complete proteome.  
SQ SEQUENCE 420 AA; 48515 MW; D142F8A6B8BAA8E CRC64;

Query Match 100.0%; Score 35; DB 17; Length 420;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
Db 392 RPPRGR 397

RESULT 10

Q9HP22 PRELIMINARY; PRT; 433 AA.  
AC Q9HP22;  
DT 01-MAR-2001 (TREMELrel. 16, Created)  
DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE VNG1842H.  
GN VNG1842H.  
OS Halobacterium sp. (strain NRC-1 / ATCC 700922 / JCM 11081).  
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;  
OC Halobacteriaceae; Halobacterium.  
CX NCBI\_TaxID=64091;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC MEDLINE=20504483; PubMed=11016950;  
RA Ng W.Y., Kennedy S.P., Mahairas G.G., Barquist B., Pan M.,  
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Shrogha J.,  
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welte R., Goo Y.A.,  
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W., Dale H.,  
RA Maddocks D.G., Jablonski P.B., Krebs M.P., Angevine C.M., Spudich J.L.,  
RA Isenbarger T.A., Peck R.P., Pohlschroder M., Dennis P.P., Omer A.D.,  
RA Alam M., Freitas T., Hou S., Daniels C.J., Riley M., Hood L., Dassarma S.;  
RA Ehardt H., Lowe T.M., Liang P., Liang P., Riley M., Hood L., Dassarma S.;  
RT "Genome sequence of Halobacterium species NRC-1."  
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181 (2000).  
DR EMBL; AE005085; AAG20048.1; -.  
DR PIR; D84335; D84335.  
KW Complete proteome.  
SQ SEQUENCE 433 AA; 48104 MW; C375590B2D15DFD0 CRC64;

Query Match 100.0%; Score 35; DB 17; Length 433;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6

Db 71 RPPRGR 76

RESULT 11

Q84SD0 PRELIMINARY; PRT; 535 AA.  
AC Q84SD0;  
DT 01-JUN-2003 (TREMELrel. 24, Created)  
DT 01-JUN-2003 (TREMELrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)  
DE P0577B11.16 protein.  
GN P0577B11.16.  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
CX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Nipponbare;  
RA Sasaki T., Matsumoto T., Katayose Y.;  
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 8, PAC  
RT clone: P0577B11.1."  
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP005504; BAC57819.1; -.  
DR InterPro; IPR002885; PPR.  
DR Pfam; PF01535; PPR; 8.  
DR TIGRFAMs; TIGR00756; PPR; 9.  
SQ SEQUENCE 535 AA; 58557 MW; C2712C4178582982 CRC64;

Query Match 100.0%; Score 35; DB 10; Length 535;  
 Best Local Similarity 100.0%; Pred. NO. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
 |||||  
 DB 115 RPPRGR 120

## RESULT 12

Q05316 PRELIMINARY; PRT; 561 AA.  
 AC Q05316;  
 DT 01-JUL-1997 (T-EMBLrel. 04, Created)  
 DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE Hypothetical protein (Diester hydrolase, putative).  
 GN RV1215C OR MTC1364.27C OR MT1253.  
 OS Mycobacterium tuberculosis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=H37RV;  
 RX MEDLINE=98285987; PubMed=9634230;  
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekait F.,  
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
 RA Hornsby R., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,  
 RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;  
 RA "Deciphering the biology of Mycobacterium tuberculosis from the  
 RT complete genome sequence."  
 RL Nature 393:537-544 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CDC 1551 / Oshkosh;  
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,  
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,  
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
 RA Bishai W.;  
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and  
 RT laboratory strains."  
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; Z93777; CAB07817.1; -;  
 DR EMBL; AE007002; AAK45510.1; -;  
 DR PIR; E70610; E70610.  
 DR TIGR; MT1253; -;  
 DR TubercuList; RV1215C; -;  
 DR GO; GO:0005634; C:nucleus; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR GO; GO:0016787; F:hydrolase activity; IEA.  
 DR GO; GO:0008451; F:xaa-pro aminopeptidase activity; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR005674; CoC NonD.  
 DR InterPro; IPR008979; Gal Bind like.  
 DR InterPro; IPR01005; Myb DNA Binding.  
 DR InterPro; IPR000383; Peptidase\_S15.  
 DR Pfam; PF02129; Peptidase\_S15; 1.  
 DR TIGRFAMs; TIGR00976; /NonD; 1.  
 DR PROSITE; PS00037; MYB 1; 1.  
 KW Hypothetical protein; Hydrolase; Complete proteome.  
 SQ SEQUENCE 561 AA; 62610 MW; 6D2C3253FD3598D CRC64;

Query Match 100.0%; Score 35; DB 16; Length 561;  
 Best Local Similarity 100.0%; Pred. NO. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
 |||||  
 DB 536 RPPRGR 541

## RESULT 13

Q7U0D5 PRELIMINARY; PRT; 561 AA.  
 AC Q7U0D5;  
 DT 01-OCT-2003 (T-EMBLrel. 25, Created)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE Conserved hypothetical protein.  
 GN MS1247C.  
 OS Mycobacterium bovis.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
 OX NCBI\_TaxID=1765;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AF2132/97;  
 RX MEDLINE=22709107; PubMed=12788972;  
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,  
 RA Pryor M., Duthey S., Gordin S., Lacroix C., Monsemp C., Simon S.,  
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,  
 RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;  
 RA "The complete genome sequence of Mycobacterium bovis."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).  
 DR EMBL; BX248338; CAD94108.1; -;  
 KW Complete proteome.  
 SQ SEQUENCE 561 AA; 62626 MW; C35176E8172866AD CRC64;

Query Match 100.0%; Score 35; DB 16; Length 561;  
 Best Local Similarity 100.0%; Pred. NO. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
 |||||  
 DB 536 RPPRGR 541

## RESULT 14

Q8S103 PRELIMINARY; PRT; 591 AA.  
 AC Q8S103;  
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)  
 DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)  
 DE P0445E10.18 protein (P0478H03.12 protein).  
 GN P0445E10.18 OR P0478H03.12.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RA "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 RT clone: P0445E10.18".  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RA "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC  
 RT clone: P0478H03.12".  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP003447; BAB86058.1; -;  
 DR EMBL; AP003452; BAB92696.1; -;  
 DR Gramene; Q8S103; -;  
 DR GO; GO:0005874; C:microtubule; IEA.



DR GO; GO:0005198; F:structural molecule activity; IEA.  
 DR GO; GO:0007018; P:microtubule-based movement; IEA.  
 DR InterPro; IPR000217; Tubulin.  
 DR PROSITE; PS00227; TUBULIN; 1.  
 SQ SEQUENCE 591 AA; 62965 MW; 6BAD242D757FD998 CRC64;

Query Match 100.0%; Score 35; DB 10; Length 591;  
 Best Local Similarity 100.0%; Pred.No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRCR 6  
 |||||  
 Db 213 RPPRCR 218

## RESULT 15

Q7Y0D7 PRELIMINARY; PRT; 658 AA.  
 AC Q7Y0D7;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein OSJNBa0079B1.5.21.  
 GN OSJNBa0079B1.5.21.  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buell C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,  
 Overton II L.L., Tsitrin T., Kim M.M., Bera J.J., Jin S.S.,  
 Fadrosh D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,  
 Vanaken S.S., Riedmuller S.B., Utterback T.T., Feldblyum T.V.,  
 Yang Q.Q., Haas B.J., Suh B.B., Peterson J.J., Quackenbush J.,  
 White O., Salzberg S.L., Fraser C.M.;  
 RT "Oryza sativa chromosome 3 BAC OSJNBa0079B1.5 genomic sequence."  
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Buell R.;  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC095043; RAP50947.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 658 AA; 72599 MW; AF255DB449072630 CRC64;

Query Match 100.0%; Score 35; DB 10; Length 658;  
 Best Local Similarity 100.0%; Pred.No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRCR 6  
 |||||  
 Db 112 RPPRCR 117

Search completed: April 6, 2004, 16:14:43  
 Job time : 22.7779 secs

GenCore version 5.1.1.6  
Copyright (C) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:29:39 ; Search time 33.5888 Seconds  
(without alignments)  
50.472 Million cell updates/sec

Title: US-10-009-709-11

Perfect score: 35

Sequence: 1 RPPRGR 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04:\*

1: Geneseq1980s:\*

2: Geneseq1990s:\*

3: Geneseq2000s:\*

4: Geneseq2001s:\*

5: Geneseq2002s:\*

6: Geneseq2003s:\*

7: Geneseq2003ss:\*

8: Geneseq2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	100.0	6	AAB48781	Aab48781 Human sal
2	35	100.0	7	AAB48780	Aab48780 Human sal
3	35	100.0	8	AAB48779	Aab48779 Human sal
4	35	100.0	9	AAB48778	Aab48778 Human sal
5	35	100.0	10	AAB48777	Aab48777 Human sal
6	35	100.0	17	AAB48783	Aab48783 Human sal
7	35	100.0	44	ABP62124	ABP62124 Human sec
8	35	100.0	63	AAU53649	AAU53649 Propionib
9	35	100.0	63	ABM50168	ABM50168 Propionib
10	35	100.0	65	AAU40624	AAU40624 Propionib
11	35	100.0	65	ABM37143	ABM37143 Propionib
12	35	100.0	69	AAU56685	AAU56685 Propionib
13	35	100.0	69	ABM53204	ABM53204 Propionib
14	35	100.0	92	ABM65798	ABM65798 Propionib
15	35	100.0	104	ABG18278	ABG18278 Novel hum
16	35	100.0	132	ABM38848	ABM38848 Peptide #
17	35	100.0	132	AAU32323	AAU32323 Peptide #
18	35	100.0	132	AAU72058	AAU72058 Human bron
19	35	100.0	132	AAU59494	AAU59494 Human bron
20	35	100.0	132	ABG53742	ABG53742 Human liv
21	35	100.0	132	ABG41873	ABG41873 Human pep
22	35	100.0	141	AAU65747	AAU65747 Propionib
23	35	100.0	141	ABM62266	ABM62266 Propionib
24	35	100.0	142	AAU57379	AAU57379 Propionib
25	35	100.0	142	ABM53898	ABM53898 Propionib

## ALIGNMENTS

### RESULT 1

AAB48781

ID AAB48781 standard; peptide; 6 AA.

AC AAB48781;

XX 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 103-108), SEQ ID NO:11.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;

KW Chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

KW oral bacterium; caries prevention.

XX Homo sapiens.

XX WO200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SE000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO//) STROEMBERG N.

XX (JOHA//) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of

XX proline-rich proteins, useful for preventing dental caries.

XX Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-AAB48783) which contain at least two arginine residues and which protect against dental caries. PRPs (proline-rich proteins) are salivary proteins encoded by six clustered genes on chromosome 12p13.2 and are potential determinants of a person's susceptibility to dental caries. PRPs are degraded by Actinomyces and Streptococcus species to small peptide fragments. These are metabolised by oral bacteria for nutritional purposes, with certain bacterial species generating ammonia via the catabolism of arginine. The peptides of the invention, being arginine-rich, can also be converted to ammonia by these bacteria. The ammonia thus formed raises the pH at the dental surface, thereby protecting the teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-

AAU42127 Propionib  
ABM38646 Propionib  
ABR57423 Human NOV  
ABR5769 Human sec  
ACD83798 Human PRH  
ACD98216 Human sal  
ABP62121 Human sec  
ABG26281 Novel hum  
ABG03818 Novel hum  
ADE07941 Novel pro  
ABG70853 Synchoco  
ABR58578 Human can  
ABR64071 Human pro  
AAB11540 SEN virus  
ABP76681 Streptomy  
AAO01213 Human pol  
AAU3535 Novel hum  
AAU32443 Novel hum  
ABO00690 Novel hum  
AAM40855 Human pol

26 35 100.0 144 4 AAU42127  
27 35 100.0 144 6 ABM38646  
28 35 100.0 149 6 ABR57423  
29 35 100.0 154 6 ABR5769  
30 35 100.0 166 6 ADA83798  
31 35 100.0 166 7 ADC98216  
32 35 100.0 182 5 ABP62121  
33 35 100.0 258 4 ABG26281  
34 35 100.0 306 4 ABG03818  
35 35 100.0 578 7 ADE07941  
36 35 100.0 694 5 ABG70853  
37 35 100.0 696 6 ABR58578  
38 35 100.0 696 7 ABR64071  
39 35 100.0 758 3 AAB11540  
40 35 100.0 19938 6 ABP76681  
41 32 91.4 91 4 AAO01213  
42 32 91.4 91 4 AAU3535  
43 32 91.4 159 4 AAU32443  
44 32 91.4 206 6 ABO00690  
45 32 91.4 285 4 AAM40855

CC derived oligopeptides of the invention

XX SQ Sequence 6 AA;

Query Match 100.0%; Score 35; DB 4; Length 6;  
Best Local Similarity 100.0%; Pred. NO. 1.4e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6

Db 1 RPPRGR 6

RESULT 2

AAAB48780  
ID AAB48780 standard; peptide; 7 AA.

XX AC AAB48780;

XX DT 09-MAR-2001 (first entry)

XX DE Human saliva PRP-1 fragment (residues 102-108), SEQ ID NO:10.

XX KW Human; PRP-1; proline-rich protein; saliva; dental caries;

XX KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;

XX KW oral bacterium; caries prevention.

XX OS Homo sapiens.

XX PN WO200069890-A1.

XX PD 23-NOV-2000.

XX PF 11-MAY-2000; 2000WO-SE000930.

XX PR 17-MAY-1999; 99SE-00001773.

XX PA (STRO/) STROEMBERG N.

XX PA (JOHA/) JOHANSSON I.

XX PI Stroemberg N, Johansson I;

XX DR WPI; 2001-031923/04.

XX PT New oligopeptides comprising 2 arginine residues from degradation of  
proline-rich proteins, useful for preventing dental caries.

XX PS Claim 4; Page 24; 36pp; English.

XX CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
AAB48783) which contain at least two arginine residues and which protect  
against dental caries. PRPs (proline-rich proteins) are salivary proteins  
encoded by six clustered genes on chromosome 12p13.2 and are potential  
determinants of a person's susceptibility to dental caries. PRPs are  
degraded by Actinomyces and Streptococcus species to small peptide  
fragments. These are metabolised by oral bacteria for nutritional  
purposes, with certain bacterial species generating ammonia via the  
catabolism of arginine. The peptides of the invention, being arginine-  
rich, can also be converted to ammonia by these bacteria. The ammonia  
thus formed raises the pH at the dental surface, thereby protecting the  
teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
derived oligopeptides of the invention

XX SQ Sequence 7 AA;

Query Match 100.0%; Score 35; DB 4; Length 7;  
Best Local Similarity 100.0%; Pred. NO. 1.4e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6

Db 2 RPPRGR 7

RESULT 3

AAAB48779  
ID AAB48779 standard; peptide; 8 AA.

XX AC AAB48779;

XX DT 09-MAR-2001 (first entry)

XX DE Human saliva PRP-1 fragment (residues 101-108), SEQ ID NO:9.

XX KW Human; PRP-1; proline-rich protein; saliva; dental caries;  
chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
oral bacterium; caries prevention.

XX OS Homo sapiens.

XX PN WO200069890-A1.

XX PD 23-NOV-2000.

XX PF 11-MAY-2000; 2000WO-SE000930.

XX PR 17-MAY-1999; 99SE-00001773.

XX PA (STRO/) STROEMBERG N.

XX PA (JOHA/) JOHANSSON I.

XX PI Stroemberg N, Johansson I;

XX DR WPI; 2001-031923/04.

XX PT New oligopeptides comprising 2 arginine residues from degradation of  
proline-rich proteins, useful for preventing dental caries.

XX PS Claim 4; Page 24; 36pp; English.

XX CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
AAB48783) which contain at least two arginine residues and which protect  
against dental caries. PRPs (proline-rich proteins) are salivary proteins  
encoded by six clustered genes on chromosome 12p13.2 and are potential  
determinants of a person's susceptibility to dental caries. PRPs are  
degraded by Actinomyces and Streptococcus species to small peptide  
fragments. These are metabolised by oral bacteria for nutritional  
purposes, with certain bacterial species generating ammonia via the  
catabolism of arginine. The peptides of the invention, being arginine-  
rich, can also be converted to ammonia by these bacteria. The ammonia  
thus formed raises the pH at the dental surface, thereby protecting the  
teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
derived oligopeptides of the invention

XX SQ Sequence 8 AA;

Query Match 100.0%; Score 35; DB 4; Length 8;  
Best Local Similarity 100.0%; Pred. NO. 1.4e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6

Db 3 RPPRGR 8

RESULT 4

AAAB48778  
ID AAB48778 standard; peptide; 9 AA.

XX AC AAB48778;

XX DT 09-MAR-2001 (first entry)

XX DE Human saliva PRP-1 fragment (residues 100-108), SEQ ID NO:8.

XX KW Human; PRP-1; proline-rich protein; saliva; dental caries;

KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.  
 XX Homo sapiens.  
 OS  
 XX WO200069890-A1.  
 XX  
 XX 23-NOV-2000.  
 XX  
 XX 11-MAY-2000; 2000WO-SE000930.  
 XX  
 XX 17-MAY-1999; 99SE-00001773.  
 XX  
 XX (STRO/) STROEMBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX  
 XX Stroemberg N, Johansson I;  
 PI  
 XX WPI; 2001-031923/04.  
 XX  
 XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 XX  
 XX Claim 4; Page 24; 36pp; English.  
 XX  
 XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention  
 XX  
 SQ Sequence 9 AA;  
 Query Match 100.0%; Score 35; DB 4; Length 9;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPPRGR 6  
 DB |||||  
 4 RPPRGR 9  
 RESULT 5  
 AAB48777  
 ID AAB48777 standard; peptide; 10 AA.  
 AC  
 XX AAB48777;  
 XX  
 XX 09-MAR-2001 (first entry)  
 DT  
 DE Human saliva PRP-1 fragment (residues 99-108), SEQ ID NO:7.  
 XX  
 XX Human; PRP-1; proline-rich protein; saliva; dental caries;  
 KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200069890-A1.  
 XX  
 XX 23-NOV-2000.  
 XX  
 XX 11-MAY-2000; 2000WO-SE000930.  
 XX  
 XX 17-MAY-1999; 99SE-00001773.  
 XX

XX (STRO/) STROEMBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX  
 XX Stroemberg N, Johansson I;  
 PI  
 XX WPI; 2001-031923/04.  
 XX  
 XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 XX  
 XX Claim 4; Page 24; 36pp; English.  
 XX  
 XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention  
 XX  
 SQ Sequence 10 AA;  
 Query Match 100.0%; Score 35; DB 4; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 22;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPPRGR 6  
 DB |||||  
 5 RPPRGR 10  
 RESULT 6  
 AAB48783  
 ID AAB48783 standard; peptide; 17 AA.  
 AC  
 XX AAB48783;  
 XX  
 XX 09-MAR-2001 (first entry)  
 DT  
 DE Human saliva PRP-1 fragment (residues 99-115), SEQ ID NO:13.  
 XX  
 XX Human; PRP-1; proline-rich protein; saliva; dental caries;  
 KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.  
 XX  
 XX Homo sapiens.  
 OS  
 XX WO200069890-A1.  
 XX  
 XX 23-NOV-2000.  
 XX  
 XX 11-MAY-2000; 2000WO-SE000930.  
 XX  
 XX 17-MAY-1999; 99SE-00001773.  
 XX  
 XX (STRO/) STROEMBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX  
 XX Stroemberg N, Johansson I;  
 PI  
 XX WPI; 2001-031923/04.  
 XX  
 XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 XX  
 XX Claim 2; Page 24; 36pp; English.  
 XX

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by *Actinomyces* and *Streptococcus* species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention  
 XX SQ Sequence 17 AA;  
 Query Match 100.0%; Score 35; DB 4; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 35;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 RPPRGR 6  
 Db |||||  
 5 RPPRGR 10  
 RESULT 7  
 ABP62124  
 ID. ABP62124 standard; protein; 44 AA.  
 XX AC  
 AC ABP62124;  
 XX DT  
 DT 12-NOV-2002. (first entry)  
 XX DE  
 DE Human secreted protein SEQ ID NO 177.  
 XX KW  
 KW Human; neutrotropic; neuroprotective; cytostatic; dermatological; virucide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
 KW antiparkinsonian; antisickling; antianemic; antiarthritic; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;  
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.  
 XX OS  
 OS Homo sapiens.  
 XX XX  
 XX WO200257420-A2.  
 XX XX  
 XX 25-JUL-2002.  
 XX XX  
 XX 17-JAN-2002; 2002WO-US001109.  
 XX PF  
 PF 18-JAN-2001; 2001US-0262066P.  
 XX PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX FA  
 XX Moore PA, Ruben SM, Lafleur DW, Shi Y, Rosen CA, Olsen H;  
 XX Eber R, Brewer LA;  
 XX WPI; 2002-599716/64.  
 XX DR  
 XX New polynucleotides and polypeptides useful for diagnosing, prognosing,  
 XX PT treating or preventing e.g. neurodegenerative, central nervous system,  
 XX PT autoimmune, respiratory, reproductive, or inflammatory diseases or  
 XX PT disorders.  
 XX PS  
 PS Claim 11; Page 54; 785pp; English.  
 XX CC  
 CC The invention relates to novel genes (ABO92553-ABO92607) and proteins  
 CC (ABP62013-ABP62153) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in

CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
 CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
 CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
 CC infectious diseases such as viral, bacterial, fungal and parasitic  
 CC infections  
 XX SQ Sequence 44 AA;  
 Query Match 100.0%; Score 35; DB 5; Length 44;  
 Best Local Similarity 100.0%; Pred. No. 79;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 RPPRGR 6  
 Db |||||  
 5 RPPRGR 10  
 RESULT 8  
 AAU53649  
 ID AAU53649 standard; protein; 63 AA.  
 XX AC  
 AC AAU53649;  
 XX DT  
 DT 27-FEB-2002 (first entry)  
 XX XX  
 DE Propionibacterium acnes immunogenic protein #14545.  
 XX KW  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX OS  
 OS Propionibacterium acnes.  
 XX PN  
 PN WO200181581-A2.  
 XX XX  
 PD 01-NOV-2001.  
 XX XX  
 PF 20-APR-2001; 2001WO-US012865.  
 XX PR  
 PR 21-APR-2000; 2000US-0199047P.  
 XX PR  
 PR 02-JUN-2000; 2000US-0208841P.  
 XX PR  
 PR 07-JUL-2000; 2000US-0216747P.  
 XX XX  
 XX (CORI-) CORIXA CORP.  
 XX XX  
 XX Skeiky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 XX L'Maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX WPI; 2001-616774/71.  
 XX DR  
 DR N-PSDB; AAS59562.  
 XX XX  
 XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 XX PT vaccinating against and diagnosing infections, especially useful for  
 XX PT treating acne vulgaris.  
 XX XX  
 XX Example 1; SEQ ID NO 14844; 1069pp; English.  
 XX PS  
 PS Sequences AAU9105-AAU68017 represent Propionibacterium acnes immunogenic  
 XX CC polypeptides. The proteins and their associated DNA sequences are used in  
 XX CC the treatment, prevention and diagnosis of medical conditions caused by  
 XX CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 XX CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 XX CC P. acnes is also involved in infections of bone, joints and the central  
 XX CC nervous system, however it is particularly involved in the inflammatory  
 XX CC lesions associated with acne vulgaris. A method for detecting the  
 XX CC presence or absence of P. acnes in a patient comprises contacting a  
 XX CC sample with a binding agent that binds to the proteins of the invention

CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 63 AA;  
 Query Match 100.0%; Score 35; DB 4; Length 63;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPPRGR 6  
 Db 20 RPPRGR 25  
 RESULT 9  
 ABM50168  
 ID ABM50168; standard; protein; 63 AA.  
 XX  
 AC ABM50168;  
 XX  
 DT 20-OCT-2003 (first entry)  
 XX  
 DE Propionibacterium acnes predicted ORF-encoded polypeptide #14844.  
 XX  
 KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
 KW immunostimulant; immune response; vaccine.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 FN WO2003033515-A1.  
 XX  
 PD 24-APR-2003.  
 XX  
 PF 11-OCT-2002; 2002WO-US032727.  
 XX  
 PR 15-OCT-2001; 2001US-00978825.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
 PI Barth B, Vallieue-Douglas J;  
 XX  
 XX WPI; 2003-381789/36.  
 DR N-PSDB; ACF64491.  
 XX  
 XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.  
 XX  
 XX Example 1; SEQ ID NO 14844; 1481bp; English.  
 XX  
 XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 CC encoding a Propionibacterium acnes protein. The invention also relates to  
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC invention; fusion proteins comprising a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising P. acnes  
 CC via this method; a vaccine composition comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide; a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a

CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridisation. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
 CC reading frame) contained within the P. acnes polynucleotides of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 63 AA;  
 Query Match 100.0%; Score 35; DB 6; Length 63;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPPRGR 6  
 Db 20 RPPRGR 25  
 RESULT 10  
 AAU40624  
 ID AAU40624 standard; protein; 65 AA.  
 XX  
 AC AAU40624;  
 XX  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #1520.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 FN WO200101581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US012865.  
 XX  
 PR 21-APR-2000; 2000US-0199047P.  
 PR 02-JUN-2000; 2000US-0208841P.  
 PR 07-JUL-2000; 2000US-0216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 PI L'Maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 XX WPI; 2001-616774/71.  
 DR N-PSDB; AAS59512.  
 XX  
 XX Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris.  
 XX  
 PS Example 1; SEQ ID NO 1819; 1069bp; English.  
 XX  
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central

CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 65 AA;  
 Query Match 100.0%; Score 35; DB 4; Length 65;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPPRGR 6  
 Db 9 RPPRGR 14  
 RESULT 11  
 ABM37143  
 ID ABM37143 standard; protein; 65 AA.  
 XX  
 AC ABM37143;  
 XX  
 DT 20-OCT-2003 (first entry)  
 XX  
 DE Propionibacterium acnes predicted ORF-encoded polypeptide #1819.  
 XX  
 KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
 KW immunostimulant; immune response; vaccine.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 FN WO2003033515-A1.  
 XX  
 PD 24-APR-2003.  
 XX  
 PF 11-OCT-2002; 2002WO-US032727.  
 XX  
 PR 15-OCT-2001; 2001US-00978825.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
 PI Barth B, Vallieue-Douglass J;  
 XX  
 DR WPI; 2003-381789/36.  
 DR N-PSDB; ACF64441.  
 XX  
 PT New Propionibacterium acnes polypeptides and polynucleotides encoding the  
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.  
 XX  
 PS Example 1; SEQ ID NO 1819; 1481bp; English.  
 XX  
 CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 CC encoding a Propionibacterium acnes protein. The invention also relates to  
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC invention; fusion proteins comprising a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising T cells prepared

CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a  
 CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridisation. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
 CC reading frame) contained within the P. acnes polynucleotides of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 65 AA;  
 Query Match 100.0%; Score 35; DB 6; Length 65;  
 Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPPRGR 6  
 Db 9 RPPRGR 14

RESULT 12  
 AAU56685  
 ID AAU56685 standard; protein; 69 AA.  
 XX  
 AC AAU56685;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #17581.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopathic; neuroprotectant.

XX Propionibacterium acnes.

XX WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US012865.

XX 21-APR-2000; 2000US-0193047P.

XX 02-JUN-2000; 2000US-0208841P.

XX 07-JUL-2000; 2000US-0216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX L'Maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI; 2001-616774/71.

XX N-PSDB; AAS59578.

XX Propionibacterium acnes polypeptides and nucleic acids useful for

XX vaccinating against and diagnosing infections, especially useful for

XX treating acne vulgaris.

XX Example 1; SEQ ID NO 17880; 1069pp; English.

XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic

XX polypeptides. The proteins and their associated DNA sequences are used in

CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central  
 CC nervous system, however it is particularly involved in the inflammatory  
 CC lesions associated with acne vulgaris. A method for detecting the  
 CC presence or absence of P. acnes in a patient comprises contacting a  
 CC sample with a binding agent that binds to the proteins of the invention  
 CC and determining the amount of bound protein in the sample. The  
 CC polypeptides may be used as antigens in the production of antibodies  
 CC specific for P. acnes proteins. These antibodies can be used to  
 CC downregulate expression and activity of P. acnes polypeptides and  
 CC therefore treat P. acnes infections. The antibodies may also be used as  
 CC diagnostic agents for determining P. acnes presence, for example, by  
 CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 69 AA;  
 SQ Query Match 100.0%; Score 35; DB 4; Length 69;  
 Best Local Similarity 100.0%; Pred. NO. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RPPRGR 6  
 DB 28 RPPRGR 33  
 |||||

RESULT 13  
 ABM53204  
 ID ABM53204 standard; protein; 69 AA.  
 XX  
 AC ABM53204;

DT 20-OCT-2003. (first entry)

DE Propionibacterium acnes predicted ORF-encoded polypeptide #17880.  
 XX  
 KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
 KW immunostimulant; immune response; vaccine.

OS Propionibacterium acnes.

XX WO2003033515-A1.

PN 24-APR-2003.

XX 11-OCT-2002; 2002WO-US032727.

XX 15-OCT-2001; 2001US-00978825.

XX (CORI-) CORIXA CORP.

PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
 PI Barth B, Vallie-Douglass J;

XX WPI; 2003-381789/36.  
 DR N-PSDB; ACF64507.

XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.

XX Example 1; SEQ ID NO 17880; 1481pp; English.

XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 CC encoding a Propionibacterium acnes protein. The invention also relates to  
 CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a

CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC invention; fusion proteins comprising a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising T cells prepared  
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a  
 CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridisation. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
 CC reading frame) contained within the P. acnes polynucleotides of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 69 AA;

Query Match 100.0%; Score 35; DB 6; Length 69;  
 Best Local Similarity 100.0%; Pred. NO. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RPPRGR 6  
 DB 28 RPPRGR 33  
 |||||

RESULT 14  
 ABM65798  
 ID ABM65798 standard; protein; 92 AA.

XX ABM65798;

DT 20-OCT-2003. (first entry)

DE Propionibacterium acnes immunogenic polypeptide #30474.

XX Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
 KW immunostimulant; immune response; vaccine; immunogenic.

OS Propionibacterium acnes.

XX WO2003033515-A1.

PN 24-APR-2003.

XX 11-OCT-2002; 2002WO-US032727.

XX 15-OCT-2001; 2001US-00978825.

XX (CORI-) CORIXA CORP.

PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
 PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
 PI Barth B, Vallie-Douglass J;

XX WPI; 2003-381789/36.

XX New Propionibacterium acnes polypeptides and polynucleotides encoding the  
 PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
 PT or for stimulating an immune response specific for a P. acnes protein.

XX Claim 7; SEQ ID NO 30474; 1481pp; English.

XX The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
 CC encoding a Propionibacterium acnes protein. The invention also relates to



CC polypeptides encoded by the polynucleotides (ABG35624-ABG64536) and to  
 CC immunogenic fragments of P. acnes polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC polynucleotide of the invention; fusion proteins comprising a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a P. acnes  
 CC polypeptide and an isolated T cell population comprising T cells prepared  
 CC via this method; a vaccine composition (comprising P. acnes polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of P. acnes in a  
 CC patient; and a method for inhibiting the development of P. acnes in a  
 CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating acne  
 CC vulgaris, or for stimulating an immune response specific for a P. acnes  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridisation. The vaccine composition is useful for the  
 CC stimulation of an immune response against P. acnes, or for treating acne,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a specifically claimed P. acnes polypeptide which is  
 CC thought to contain an immunogenic region. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX

SQ Sequence 92 AA;

Query Match 100.0%; Score 35; DB 6; Length 92;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
 |||||  
 Db 18 RPPRGR 23

RESULT 15  
 ABG18278  
 ID ABG18278 standard; protein; 104 AA.

AC ABG18278;

DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #18269.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

XX WC200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

PR 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS82465.

XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity.

XX

PS Claim 20; SEQ ID NO 48637; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have application in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX

SQ Sequence 104 AA;

Query Match 100.0%; Score 35; DB 4; Length 104;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
 |||||  
 Db 94 RPPRGR 99

Search completed: April 6, 2004, 16:06:41

Job time : 33.5888 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 16:14:50 ; Search time 23.5514 Seconds  
(without alignments)  
66.909 Million cell updates/sec

Title: US-10-009-709-11

Perfect score: 35

Sequence: 1 RPPRGR 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 26263353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/FCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	100.0	44	US-10-411-224-177	Sequence 177, App
2	35	100.0	44	US-10-047-021-177	Sequence 177, App
3	35	100.0	72	US-10-424-599-223703	Sequence 223703
4	35	100.0	103	US-10-425-114-40547	Sequence 40547, A
5	35	100.0	111	US-10-424-599-163482	Sequence 163482, A
6	35	100.0	129	US-10-425-114-51211	Sequence 51211, A
7	35	100.0	132	US-09-864-761-43644	Sequence 43644, A
8	35	100.0	166	US-10-157-031-80	Sequence 80, Appl
9	35	100.0	182	US-10-411-224-174	Sequence 174, App
10	35	100.0	182	US-10-047-021-174	Sequence 174, App
11	35	100.0	186	US-10-156-761-11296	Sequence 11296, A
12	35	100.0	212	US-10-425-114-65970	Sequence 65970, A
13	35	100.0	267	US-10-156-761-10080	Sequence 10080, A
14	35	100.0	268	US-10-425-114-65233	Sequence 65233, A
15	35	100.0	279	US-10-425-114-63981	Sequence 63981, A

15	35	100.0	302	12	US-10-425-114-55640	Sequence 55640, A
17	35	100.0	316	12	US-10-425-114-57324	Sequence 57324, A
18	35	100.0	368	12	US-10-425-114-57774	Sequence 57774, A
19	35	100.0	395	12	US-10-425-114-55129	Sequence 55129, A
20	35	100.0	396	12	US-10-425-114-57494	Sequence 57494, A
21	35	100.0	428	12	US-10-425-114-42505	Sequence 42505, A
22	35	100.0	477	12	US-10-425-114-71670	Sequence 71670, A
23	35	100.0	696	15	US-10-104-047-22325	Sequence 22325, App
24	35	100.0	19652	15	US-10-084-846A-7	Sequence 7, Appli
25	32	91.4	71	12	US-10-424-599-178111	Sequence 178111, A
26	32	91.4	111	12	US-10-424-599-177430	Sequence 177430, A
27	32	91.4	164	12	US-10-424-599-174830	Sequence 174830, A
28	32	91.4	314	9	US-09-768-826-43	Sequence 43, Appl
29	32	91.4	314	11	US-09-833-245-2266	Sequence 2266, App
30	32	91.4	326	12	US-10-276-774-2486	Sequence 2486, App
31	32	91.4	518	9	US-09-925-299-862	Sequence 862, App
32	32	91.4	518	10	US-09-925-299-862	Sequence 862, App
33	32	91.4	555	14	US-10-106-698-5847	Sequence 5847, App
34	32	91.4	587	14	US-10-259-165-134	Sequence 134, App
35	32	91.4	587	14	US-10-259-165-464	Sequence 464, App
36	32	91.4	667	10	US-09-863-776-16	Sequence 16, Appl
37	32	91.4	667	10	US-09-863-776-18	Sequence 18, Appl
38	32	91.4	669	10	US-09-823-187-88	Sequence 88, Appl
39	32	91.4	669	10	US-09-863-776-53	Sequence 53, Appl
40	32	91.4	669	10	US-09-863-776-54	Sequence 54, Appl
41	32	91.4	670	10	US-09-823-187-86	Sequence 86, Appl
42	32	91.4	670	10	US-09-863-776-51	Sequence 51, Appl
43	32	91.4	671	10	US-09-946-374-308	Sequence 308, App
44	32	91.4	671	10	US-09-823-187-85	Sequence 85, Appl
45	32	91.4	671	10	US-09-863-776-20	Sequence 20, Appl

#### ALIGNMENTS

#### RESULT 1

US-10-411-224-177  
; Sequence 177, Application US/10411224  
; Publication No. US20030166906A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 50 Human Secreted Proteins  
; FILE REFERENCE: P2016P1  
; CURRENT APPLICATION NUMBER: US/10/411,224  
; CURRENT FILING DATE: 2003-04-11  
; PRIOR APPLICATION NUMBER: US/09/722,329  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: 09/262,109  
; PRIOR FILING DATE: 1999-03-04  
; PRIOR APPLICATION NUMBER: 60/057,626  
; PRIOR FILING DATE: 1997-09-05  
; PRIOR APPLICATION NUMBER: 60/057,663  
; PRIOR FILING DATE: 1997-09-05  
; PRIOR APPLICATION NUMBER: 60/057,669  
; PRIOR FILING DATE: 1997-09-05  
; PRIOR APPLICATION NUMBER: 60/058,667  
; PRIOR FILING DATE: 1997-09-12  
; PRIOR APPLICATION NUMBER: 60/058,974  
; PRIOR FILING DATE: 1997-09-12  
; PRIOR APPLICATION NUMBER: 60/058,973  
; PRIOR FILING DATE: 1997-09-12  
; PRIOR APPLICATION NUMBER: 60/058,666  
; PRIOR FILING DATE: 1997-09-12  
; PRIOR APPLICATION NUMBER: 60/090,112  
; PRIOR FILING DATE: 1998-06-22  
; NUMBER OF SEQ ID NOS: 206  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 177  
; LENGTH: 44  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-411-224-177

```
Query Match      100.0%; Score 35; DB 14; Length 44;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6
Db 5 RPPRGR 10

RESULT 2
US-10-047-021-177
; Sequence 177, Application US/10047021
; Publication No. US20040002591A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 50 Human Secreted Proteins
; FILE REFERENCE: P2016P2
; CURRENT APPLICATION NUMBER: US/10/047,021
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/262,066
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 09/722,329
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US 09/262,109
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: PCT/US98/18360
; PRIOR FILING DATE: 1998-09-03
; PRIOR APPLICATION NUMBER: US 60/057,626
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 60/057,663
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 60/057,669
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 60/058,667
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: US 60/058,974
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: US 60/058,973
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: US 60/058,666
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: US 60/090,112
; PRIOR FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 177
; LENGTH: 44
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-047-021-177

Query Match      100.0%; Score 35; DB 15; Length 44;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6
Db 5 RPPRGR 10

RESULT 3
US-10-424-599-223703
; Sequence 223703, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 163482
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
US-10-424-599-223703
```

```
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 223703
; LENGTH: 72
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_44033C.1.pap
US-10-424-599-223703

Query Match      100.0%; Score 35; DB 12; Length 72;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6
Db 28 RPPRGR 33

RESULT 4
US-10-425-114-40547
; Sequence 40547, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E.
; APPLICANT: Tabaska, Jack E.
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53213)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 40547
; LENGTH: 103
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB189-001-G7_FUJ.pap
US-10-425-114-40547

Query Match      100.0%; Score 35; DB 12; Length 103;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6
Db 77 RPPRGR 82

RESULT 5
US-10-424-599-163482
; Sequence 163482, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 163482
; LENGTH: 111
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
US-10-424-599-163482
```

```

; NAME/KEY: unsure
; LOCATION: (1)..(111)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_WRT3947_118643C.1.pap
US-10-424-599-163482

Query Match      100.0%; Score 35; DB 12; Length 111;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RPPRGR 6
Db      15 RPPRGR 20

RESULT 6
US-10-425-114-51211
; Sequence 51211, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 51211
; LENGTH: 129
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700615179_FLI.pap
US-10-425-114-51211

Query Match      100.0%; Score 35; DB 12; Length 129;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RPPRGR 6
Db      106 RPPRGR 111

RESULT 7
US-09-864-761-43644
; Sequence 43644, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR FILING DATE: 2001-05-23
; PRIOR FILING DATE: 2000-02-04
; PRIOR FILING DATE: 2000-02-04
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-05-26
; PRIOR FILING DATE: 2000-08-03
; PRIOR FILING DATE: 2000-08-03
; PRIOR FILING DATE: 2000-10-04
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359

```

```

; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 43644
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC006518.17
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77
; OTHER INFORMATION: EST HUMAN HIT: BF088785.1, EVALUE 1.00e-01
; OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02
US-09-864-761-43644

Query Match      100.0%; Score 35; DB 9; Length 132;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RPPRGR 6
Db      85 RPPRGR 90

RESULT 8
US-10-157-031-80
; Sequence 80, Application US/10157031
; Publication No. US2003010890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Kravkovskaya, L. L.
; TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences
; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80
; LENGTH: 166
; TYPE: PRT

```

Tue Apr 6 17:16:01 2004

us-10-009-709-11.rapb

```
; ORGANISM: Homo sapiens
US-10-157-031-80

Query Match      100.0%; Score 35; DB 14; Length 166;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6
DB 119 RPPRGR 124

RESULT 9
US-10-411-224-174
; Sequence 174, Application US/10411224
; Publication No. US20030166906A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 50 Human Secreted Proteins
; FILE REFERENCE: P2016P1
; CURRENT APPLICATION NUMBER: US/10/411,224
; CURRENT FILING DATE: 2003-04-11
; PRIOR APPLICATION NUMBER: US/09/722,329
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 09/262,109
; PRIOR FILING DATE: 1999-03-04
; PRIOR APPLICATION NUMBER: 60/057,626
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/057,663
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/057,669
; PRIOR FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: 60/058,667
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: 60/058,974
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: 60/058,973
; PRIOR FILING DATE: 1997-09-12
; PRIOR APPLICATION NUMBER: 60/058,666
; PRIOR FILING DATE: 1998-06-22
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 174
; LENGTH: 182
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-411-224-174

Query Match      100.0%; Score 35; DB 14; Length 182;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6
DB 90 RPPRGR 95

RESULT 10
US-10-411-224-174
; Sequence 174, Application US/10047021
; Publication No. US20040002591A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 50 Human Secreted Proteins
; FILE REFERENCE: P2016P2
; CURRENT APPLICATION NUMBER: US/10/047,021
; CURRENT FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: US 60/262,066
; PRIOR FILING DATE: 2001-01-18
; PRIOR APPLICATION NUMBER: US 09/722,329
; PRIOR FILING DATE: 2000-11-28

Query Match      100.0%; Score 35; DB 14; Length 186;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6
DB 12 RPPRGR 17

RESULT 11
US-10-156-761-11296
; Sequence 11296, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11296
; LENGTH: 186
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11296

Query Match      100.0%; Score 35; DB 14; Length 186;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6
DB 12 RPPRGR 17
```

```
RESULT 12
US-10-425-114-65770
; Sequence 65770, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Zhou, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65770
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3150-022-D11_FLI.pep
US-10-425-114-65770
Query Match 100.0%; Score 35; DB 12; Length 212;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6
DB 8 RPPRGR 13

RESULT 13
US-10-156-761-10080
; Sequence 10080, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10080
; LENGTH: 267
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10080
Query Match 100.0%; Score 35; DB 14; Length 267;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6
DB 167 RPPRGR 172

RESULT 14
US-10-425-114-65233
; Sequence 65233, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65233
; LENGTH: 268
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4744-011-F5_FLI.pep
US-10-425-114-65233
Query Match 100.0%; Score 35; DB 12; Length 268;
Best Local Similarity 100.0%; Pred. No. 3.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6
DB 16 RPPRGR 21

RESULT 15
US-10-425-114-63981
; Sequence 63981, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 63981
; LENGTH: 279
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3245-070-F9_FLI.pep
US-10-425-114-63981
Query Match 100.0%; Score 35; DB 12; Length 279;
Best Local Similarity 100.0%; Pred. No. 3.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6
DB 209 RPPRGR 214

Search completed: April 6, 2004, 17:06:08
Job time : 24.5514 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:56:34 ; Search time 8.80374 Seconds  
(without alignments)  
35.185 Million cell updates/sec

Title: US-10-009-709-11

Perfect score: 35

Sequence: 1 RPPRGR 6

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A-COMB.pep:\*

2: /cgn2\_6/ptodata/2/iaa/5S-COMB.pep:\*

3: /cgn2\_6/ptodata/2/iaa/6A-COMB.pep:\*

4: /cgn2\_6/ptodata/2/iaa/6B-COMB.pep:\*

5: /cgn2\_6/ptodata/2/iaa/PCITUS-COMB.pep:\*

6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	35	100.0	175	4	US-09-252-991A-28806
2	35	100.0	177	4	US-09-252-991A-29848
3	35	100.0	195	4	US-09-252-991A-21451
4	35	100.0	207	4	US-09-252-991A-29505
5	35	100.0	223	4	US-09-252-991A-17072
6	35	100.0	302	4	US-09-252-991A-19798
7	35	100.0	334	4	US-09-252-991A-22532
8	35	100.0	418	4	US-09-252-991A-29452
9	35	100.0	559	4	US-09-252-991A-24480
10	35	100.0	748	4	US-09-252-991A-28449
11	35	100.0	1228	4	US-09-252-991A-17764
12	32	91.4	295	4	US-09-252-991A-28897
13	32	91.4	1433	2	US-08-365-486A-21
14	32	91.4	1433	3	US-09-123-708-4
15	32	91.4	1433	3	US-08-880-342-4
16	32	91.4	1433	3	US-08-880-342-21
17	32	91.4	1434	3	US-08-880-342-19
18	32	91.4	1434	3	US-08-880-342-19
19	32	91.4	1434	2	US-08-661-258-1
20	32	91.4	1554	2	US-08-705-625-3
21	32	91.4	1554	3	US-09-010-998-6
22	32	91.4	1554	3	US-09-220-574-3
23	31	88.6	240	4	US-09-252-991A-28112
24	31	88.6	358	4	US-09-252-991A-20171
25	31	88.6	377	4	US-09-252-991A-29332
26	31	88.6	422	4	US-08-403-852D-17
27	31	88.6	422	3	US-08-510-6465-18

## ALIGNMENTS

### RESULT 1

US-09-252-991A-28806

; Sequence 28806, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 28806

; LENGTH: 175

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-28806

Query Match 100.0%; Score 35; DB 4; Length 175;

Best Local Similarity 100.0%; Pred. No. 43;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 RPPRGR 6

Db 126 RPPRGR 131

### RESULT 2

US-09-252-991A-29848

; Sequence 29848, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 29848

; LENGTH: 177

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29848

Query Match  
Best Local Similarity 100.0%; Score 35; DB 4; Length 177;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPPRGR 6  
Db 47 RPPRGR 52

## RESULT 3

US-09-252-991A-21451  
; Sequence 21451, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21451  
; LENGTH: 195  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
; FEATURE:  
; NAME/KEY: UNSURE  
; LOCATION: (128)  
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.  
US-09-252-991A-21451

Query Match  
Best Local Similarity 100.0%; Score 35; DB 4; Length 195;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPPRGR 6  
Db 71 RPPRGR 75

## RESULT 4

US-09-252-991A-29505  
; Sequence 29505, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 29505  
; LENGTH: 207  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29505

Query Match  
Best Local Similarity 100.0%; Score 35; DB 4; Length 207;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPPRGR 6  
Db 85 RPPRGR 90

## RESULT 5

US-09-252-991A-17072  
; Sequence 17072, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 17072  
; LENGTH: 223  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17072

Query Match  
Best Local Similarity 100.0%; Score 35; DB 4; Length 223;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPPRGR 6  
Db 141 RPPRGR 146

## RESULT 6

US-09-252-991A-19798  
; Sequence 19798, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 19798  
; LENGTH: 302  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-19798

Query Match  
Best Local Similarity 100.0%; Score 35; DB 4; Length 302;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPPRGR 6  
Db 64 RPPRGR 69

## RESULT 7

US-09-252-991A-22532  
; Sequence 22532, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.



```
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 22532
/ LENGTH: 334
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22532

Query Match      100.0%; Score 35; DB 4; Length 334;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RPPRGR 6
Db      127 RPPRGR 132

RESULT 8
US-09-252-991A-29452
/ Sequence 29452, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 29452
/ LENGTH: 418
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29452

Query Match      100.0%; Score 35; DB 4; Length 418;
Best Local Similarity 100.0%; Pred. No. 94;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RPPRGR 6
Db      162 RPPRGR 167

RESULT 9
US-09-252-991A-24480
/ Sequence 24480, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 24480
```

```
/ LENGTH: 559
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-24480

Query Match      100.0%; Score 35; DB 4; Length 559;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RPPRGR 6
Db      316 RPPRGR 321

RESULT 10
US-09-252-991A-28449
/ Sequence 28449, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 28449
/ LENGTH: 748
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-28449

Query Match      100.0%; Score 35; DB 4; Length 748;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 RPPRGR 6
Db      114 RPPRGR 119

RESULT 11
US-09-252-991A-17764
/ Sequence 17764, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 17764
/ LENGTH: 1228
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
/ FEATURE:
/ NAME/KEY: UNSURE
/ LOCATION: (17)
/ OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-252-991A-17764

Query Match      100.0%; Score 35; DB 4; Length 1228;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
```

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
|||||

Db 1000 RPPRGR 1005

RESULT 12  
US-09-252-991A-28997  
; Sequence 28997, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 28957  
; LENGTH: 295  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-28997

Query Match 91.4%; Score 32; DB 4; Length 295;  
Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
|||||

Db 42 RPPRGR 47

RESULT 13  
US-08-365-486A-21  
; Sequence 21, Application US/08365486A  
; Patent No. 5834306  
; GENERAL INFORMATION:  
; APPLICANT: Webster, Keith A.  
; APPLICANT: Bishopric, Nanette H.  
; TITLE OF INVENTION: Tissue Specific Hypoxia Regulated  
; TITLE OF INVENTION: Therapeutic Constructs  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dehlinger & Associates  
; STREET: 350 Cambridge Avenue, Suite 250  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/365,486A  
; FILING DATE: 23-DEC-1994  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sholtz, Charles K.  
; REGISTRATION NUMBER: 38,615  
; REFERENCE/DOCKET NUMBER: 8255-0018  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 324-0880  
; TELEFAX: (415) 324-0960  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 1433 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-365-486A-21

Query Match 91.4%; Score 32; DB 2; Length 1433;  
Best Local Similarity 83.3%; Pred. No. 9.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
|||||

Db 515 KPPRGR 520

RESULT 14  
US-09-123-708-4  
; Sequence 4, Application US/09123708  
; Patent No. 6146887  
; GENERAL INFORMATION:  
; APPLICANT: SCHRADER, Jurgen  
; APPLICANT: GORDECKE, Axel  
; TITLE OF INVENTION: DNA EXPRESSION VECTORS FOR USE IN GENE THERAPEUTIC  
; FILE REFERENCE: 511169-2003  
; CURRENT APPLICATION NUMBER: US/09/123,708  
; CURRENT FILING DATE: 1998-07-28  
; EARLIER APPLICATION NUMBER: 08/553,503  
; EARLIER FILING DATE: 1996-03-01  
; EARLIER APPLICATION NUMBER: P4411402.8  
; EARLIER FILING DATE: 1994-03-31  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 4  
; LENGTH: 1433  
; TYPE: PRT  
; ORGANISM: Cytomegalovirus  
US-09-123-708-4

Query Match 91.4%; Score 32; DB 3; Length 1433;  
Best Local Similarity 83.3%; Pred. No. 9.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
|||||

Db 515 KPPRGR 520

RESULT 15  
US-09-123-624-4  
; Sequence 4, Application US/09123624  
; Patent No. 6149936  
; GENERAL INFORMATION:  
; APPLICANT: SCHRADER, Jurgen  
; APPLICANT: GORDECKE, Axel  
; TITLE OF INVENTION: DNA EXPRESSION VECTORS FOR USE IN THE GENE THERAPEUTIC  
; FILE REFERENCE: 511169-2004  
; CURRENT APPLICATION NUMBER: US/09/123,624  
; CURRENT FILING DATE: 1998-07-28  
; PRIOR APPLICATION NUMBER: 08/553,503  
; PRIOR FILING DATE: 1996-03-01  
; PRIOR APPLICATION NUMBER: 4411402.8  
; PRIOR FILING DATE: 1994-03-31  
; NUMBER OF SEQ ID NOS: 6  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 1433  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-123-624-4

Query Match 91.4%; Score 32; DB 3; Length 1433;

Best Local Similarity 83.3%; Pred. No. 9.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6

Db :|||||  
515 KPPRGR 520

Search completed: April 6, 2004, 16:19:41  
Job time : 8.80374 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:31:19 ; Search time 3.86916 Seconds  
(without alignments)  
80.746 Million cell updates/sec

Title: US-10-009-709-11

Perfect score: 35

Sequence: 1 RPPRGR 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Match Length	DB ID	Description
1	35	100.0	166	1	P02810 homo sapien
2	35	100.0	337	1	P52562 halobacteri
3	32	91.4	486	1	O95704 homo sapien
4	32	91.4	504	1	O35827 rattus norv
5	32	91.4	504	1	P55501 rhizobium s
6	32	91.4	576	1	Q8t168 homo sapien
7	32	91.4	579	1	Q9eqj4 rattus norv
8	32	91.4	614	1	Q10600 m cyns/cyrc
9	32	91.4	1429	1	P29475 homo sapien
10	32	91.4	1434	1	P29475 homo sapien
11	31	88.6	260	1	O70152 mus musculu
12	31	88.6	266	1	Q9wa83 cricetus
13	31	88.6	342	1	Q43349 arabidopsis
14	31	88.6	421	1	P54991 streptomyc
15	31	88.6	888	1	Q8r4p4 mus musculu
16	30	85.7	101	1	Q9eqj4 rattus norv
17	30	85.7	108	1	P27553 human papil
18	30	85.7	137	1	P05352 agrobacteri
19	30	85.7	141	1	P71706 mycobacteri
20	30	85.7	198	1	Q44622 homo sapien
21	30	85.7	202	1	Q49626 mycobacteri
22	30	85.7	237	1	Q9byl1 homo sapien
23	30	85.7	256	1	O50571 streptomyc
24	30	85.7	263	1	P04600 human immu
25	30	85.7	265	1	O34928 bacillus su
26	30	85.7	272	1	Q04729 bacillus st
27	30	85.7	278	1	Q02277 mycobacteri
28	30	85.7	279	1	P09255 varicella-z
29	30	85.7	292	1	Q9a725 caulobacter
30	30	85.7	305	1	Q9d338 mus musculu
31	30	85.7	335	1	P55389 rhizobium s
32	30	85.7	340	1	P41358 rhinderpest
33	30	85.7	350	1	P09891 arthrobacte
34	30	85.7	469	1	P07646 pseudorabie

Q9wud6 mus musculu  
O14904 homo sapien  
C8r5m2 mus musculu  
Q22289 caenorhabdi  
P03230 epstein-bar  
P13198 epstein-bar  
Q9z014 mus musculu  
P03345 human t-cel  
P14076 human t-cel  
P14077 human t-cel  
Q9yeb2 aeropyrum p  
Q9h3z7 homo sapien

34 30 85.7 350 1 WN8B\_MOUSE  
35 30 85.7 365 1 WN9A\_HUMAN  
36 30 85.7 365 1 WN9A\_MOUSE  
37 30 85.7 377 1 TBX9\_CABEL  
38 30 85.7 386 1 LMP1\_EBV  
39 30 85.7 386 1 LMP1\_EBVR  
40 30 85.7 392 1 EM15\_MOUSE  
41 30 85.7 429 1 GAG\_HTLIA  
42 30 85.7 429 1 GAG\_HTLIC  
43 30 85.7 429 1 GAG\_HTLIM  
44 30 85.7 438 1 SYH\_AERPE  
45 30 85.7 469 1 CTD5\_HUMAN

#### ALIGNMENTS

#### RESULT 1

PRPC\_HUMAN STANDARD; PRT; 166 AA.  
ID PRPC\_HUMAN  
AC P02810;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-3) (PRP-2/PRP-4) (PIF-F/PIF-S) (Protein A/protein C) [Contains: Peptide P-C].  
DE PRP1 AND PRH2.  
GN Homo sapiens (Human).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ALLELES PRH1-4 AND PRH2-1).  
RX MEDLINE=86196106; PubMed=3009472;  
RA Kim H.-S., Maeda N.;  
RT "Structures of two Haellii-type genes in the human salivary proline-rich protein multigene family.";  
RL J. Biol. Chem. 261:6712-6718(1986).  
RN [2]  
RP SEQUENCE FROM N.A. (ALLELES PRH2-1 AND PRH1-PIF).  
RX MEDLINE=85289325; PubMed=2993301;  
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
RT "Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system.";  
RL J. Biol. Chem. 260:11123-11130(1985).  
RN [3]  
RP SEQUENCE OF 17-166 (PRP-1 TO PRP-4; PIF-F AND PIF-S).  
RX MEDLINE=89061650; PubMed=3196309;  
RA Hay D.I., Bennick A., Schlesinger D.H., Minaguchi K., Madapallimattam G., Schluckebier S.K.;  
RT "The primary structures of six human salivary acidic proline-rich proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-f).";  
RL Biochem. J. 255:15-21(1988).  
RN [4]  
RP SEQUENCE OF 17-166 FROM N.A. (PA ALLELE).  
RX MEDLINE=88074309; PubMed=3687941;  
RA Azen E.A., Kim H.-S., Goodman P., Flynn S., Maeda N.;  
RT "Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich proteins Pa Db and PIF.";  
RL Am. J. Hum. Genet. 41:1035-1047(1987).  
RN [5]  
RP SEQUENCE OF 17-166 (PRP-2).  
RX MEDLINE=86222916; PubMed=3710693;  
RA Schlesinger D.H., Hay D.I.;  
RT "Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of calcium phosphate crystal growth from human parotid saliva.";  
RL Int. J. Pept. Protein Res. 27:373-379(1986).  
RN [6]  
RP SEQUENCE OF 17-166 (PROTEIN C).  
RX MEDLINE=80204368; PubMed=7380845;  
RA Wong R.S.C., Bennick A.;

RT "The primary structure of a salivary calcium-binding proline-rich phosphoprotein (protein C), a possible precursor of a related salivary protein A." J. Biol. Chem. 255:5943-5948 (1980).

[7] SEQUENCE OF 17-46 (PROTEIN C). MEDLINE=81191179; PubMed=7228490;

RA Schlesinger D.H., Hay D.I.;

RT "Primary structure of the active tryptic fragments of human and monkey salivary anionic proline-rich proteins." Int. J. Pept. Protein Res. 17:34-41 (1981).

[8] SEQUENCE OF 17-122 (PROTEIN A). MEDLINE=79173237; PubMed=438215;

RA Wong R.S.C., Hofmann T., Bennick A.;

RT "The complete primary structure of a proline-rich phosphoprotein from human saliva." J. Biol. Chem. 254:4800-4808 (1979).

[9] SEQUENCE OF 17-122 (PROTEIN A). Schlesinger D.H., Hay D.I.;

RA "Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent inhibitor of calcium phosphate precipitation in human parotid saliva." (In) Gross E., Meienhofer J. (eds.);

RL Peptides: structure and biological function (Proceedings of the 6th American peptide symposium), pp.133-136, Pierce Chemical Co., Rockford IL. (1979).

[10] SEQUENCE OF 123-166 (PEPTIDE P-C). MEDLINE=80227634; PubMed=7390979;

RA Isemura S., Saitoh E., Sanada K.;

RT "The amino acid sequence of a salivary proline-rich peptide, P-C, and its relation to a salivary proline-rich phosphoprotein, protein C." J. Biochem. 87:1071-1077 (1980).

[11] VARIANT PRH2-3 LYS-163. Azen E.A.;

RA "A frequent mutation in the acidic proline-rich protein gene, PRH2, causing a Q147K change closely adjacent to the bacterial binding domain of the cognate salivary PRP (Pri') in Afro-Americans." Hum. Mutat. 12:72-72 (1998).

CC -!- FUNCTION: PRP's act as highly potent inhibitors of crystal growth of calcium phosphates. They provide a protective and reparative environment for dental enamel which is important for the integrity of the teeth.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- PTM: Proteolytically cleaved; PRP-2, PRP-1, and PIF-S yield PRP-4, PRP-3 (protein B), and PIF-F, respectively.

CC -!- POLYMORPHISM: Allele PRH1-4 is also known as PRH1 DA allele; allele PRH2-1 is also known as PR1 or protein C; allele PRH2-3 is also known as PR1'.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; K03202; AAA60183.1; -

DR EMBL; K03203; AAA60184.1; -

DR EMBL; M13057; AAA98807.1; -

DR EMBL; M13058; AAA98808.1; -

DR Genew; HGNC:9366; PRH1.

DR Genew; HGNC:9367; PRH2.

DR MIM; 168730; -

DR MIM; 168790; -

DR MIM; 168710; -

DR GO; GO:0005615; Extracellular space; TAS.

DR Repeat; Parotid gland; Phosphorylation; Signal; Polymorphism;

KW Pyrrolidone carboxylic acid.

FT SIGNAL 1 16 SALIVARY ACIDIC PROLINE-RICH

FT CHAIN 17 166 PHOSPHOPROTEIN 1/2.

FT CHAIN 17 122 SALIVARY ACIDIC PROLINE-RICH

FT CHAIN 123 166 PHOSPHOPROTEIN 3/4.

FT DOMAIN 17 46 PEPTIDE P-C.

FT MOD\_RES 17 17 INHIBIT HYDROXYAPATITE FORMATION, BIND

FT MOD\_RES 24 24 TO HYDROXYAPATITE AND CALCIUM.

FT MOD\_RES 38 38 PYRROLIDONE CARBOXYLIC ACID.

FT VARIAT 20 20 PHOSPHORYLATION.

FT VARIAT 66 66 /FTID=VAR\_005563.

FT VARIAT 163 163 /FTID=VAR\_005564.

FT CONFLICT 41 41 /FTID=VAR\_005565.

FT CONFLICT 41 41 F -> P (IN REF. 10).

SQ SEQUENCE 166 AA; 17017 MW; A7DF62BF94E3C3EF CRC64;

Query Match 100.0%; Score 35; DB 1; Length 166;

Best Local Similarity 100.0%; Pred. No. 7.8;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPRGR 6

Db 119 RPRGR 124

RESULT 2

TRPD\_HALVO STANDARD; PRT; 337 AA.

ID TRPD\_HALVO STANDARD; PRT; 337 AA.

AC P2562;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Anthranilate phosphoribosyltransferase (EC 2.4.2.18).

GN TRPD.

OS Halobacterium volcanii (Haloflex volcanii).

OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;

OC Halobacteriaceae; Haloflex.

OX NCBI\_TaxID=2246;

[1] RN

RP SEQUENCE FROM N.A.

RC STRAIN=WF11;

RX MEDLINE=92165748; PubMed=1537810;

RA Lam W.L., Logan S.M., Doolittle W.F.;

RT "Genes for tryptophan biosynthesis in the halophilic archaeobacterium Haloflex volcanii: the trpPFG cluster." J. Bacteriol. 174:1694-1697 (1992).

RL -!- CATALYTIC ACTIVITY: Anthranilate + phosphoribosyl-diphosphate = N-5'-phosphoribosyl-anthranilate + diphosphate.

CC -!- PATHWAY: Tryptophan biosynthesis; second step.

CC -!- SIMILARITY: Belongs to the anthranilate phosphoribosyltransferase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M83788; AAA73175.1; -

DR HANAP; MF\_00211; -

DR InterPro; IPR005940; Ant\_phospho\_trans.

DR InterPro; IPR000312; Glyco\_trans\_3.

DR Pfam; PF02885; Glycos\_trans\_3N; 1.

DR Pfam; PF00591; Glycos\_transf\_3; 1.

DR ProDom; PD001864; Glyco\_trans\_3; 1.

DR TIGRFAMs; TIGR01245; trpd; 1.

DR Tryptophan biosynthesis; Transferase; Glycosyltransferase.

KW

[illegible]

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; Y13413; CAA73837.1; -  
 DR InterPro; IPR006020; PTB.PID.  
 DR InterPro; IPR001202; WW\_RspS\_WWP.  
 DR Pfam; PF00640; PID; 2.  
 DR Pfam; PF00397; WW; 1.  
 DR SMART; SM00462; PTB; 2.  
 DR SMART; SM00456; WW; 1.  
 DR PROSITE; PS01179; PID; 2.  
 DR PROSITE; PS01159; WW\_DOMAIN\_1; 1.  
 DR PROSITE; PS0020; WW\_DOMAIN\_2; 1.  
 KW Repeat.  
 FT DOMAIN 29 61 WW.  
 FT DOMAIN 117 252 PID 1.  
 FT DOMAIN 288 413 PID 2.  
 SQ SEQUENCE 504 AA; 54907 MW; A49CA9E99BDB3A3A CRC64;

Query Match 91.4%; Score 32; DB 1; Length 504;  
 Best Local Similarity 83.3%; Pred. No. 86;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6  
 DB 78 RPPKGR 83

## RESULT 5

Y4JA\_RHISN STANDARD; PRT; 504 AA.  
 ID Y4JA\_RHISN  
 AC P55501;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DT Hypochemical 57.2 kDa protein Y4JA/Y4NE/Y4SE.  
 GN Y4JA AND Y4NE AND Y4SE.  
 OS Rhizobium sp. (strain NGR234).  
 OG Plasmid sym pNGR234.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
 OX NCBI\_TaxID=394;  
 [1]

SEQUENCE FROM N.A.  
 RX MEDLINE=97305956; PubMed=9163424;  
 RA Freiberg C.A., Feilay R., Bairoch A., Broughton W.J., Rosenthal A.,  
 RA Perret X.;  
 RT "Molecular basis of symbiosis between Rhizobium and legumes."  
 RL Nature 387:394-401(1997).  
 CC -!- SIMILARITY: VERY LOW SIMILARITY TO THE IS21/IS408/IS1162 FAMILY OF  
 CC TRANSPOSASES.  
 CC -!- SIMILARITY: STRONG, TO PZ4.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; A8000079; AAB91713.1; -  
 DR EMBL; A8000086; AAB91785.1; -  
 DR EMBL; A8000095; AAB91845.1; -  
 DR InterPro; IPR001584; Rve.  
 DR Pfam; PF00665; rve; 1.  
 KW Hypochemical protein; Plasmid.  
 SQ SEQUENCE 504 AA; 57228 MW; 63E31E487DD4CB87 CRC64;

Query Match 91.4%; Score 32; DB 1; Length 504;  
 Best Local Similarity 83.3%; Pred. No. 86;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RPPRGR 6  
 DB 51 KPFRGR 56

## RESULT 6

Z384\_HUMAN STANDARD; PRT; 576 AA.  
 ID Z384\_HUMAN  
 AC Q8TF68; O15407; Q8N938;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Zinc finger protein 384 (Nuclear matrix transcription factor 4)  
 DE (CAG repeat protein 1).  
 OS Homo sapiens (Human).  
 GN ZNF384 OR NMF4 OR CAGH1.  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RA Matsuo M.Y.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Brain;  
 RA Tanigami A., Fujiwara T., Shibahara T., Goto Y., Hirao M., Shimizu F.,  
 RA Wakebe H., Ono T., Hishigaki H., Watanabe T., Ozaki K., Sugiyama T.,  
 RA Irie R., Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J.,  
 RA Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,  
 RA Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,  
 RA Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,  
 RA Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K.,  
 RA Masuno Y., Negai K., Isogai T.;  
 RA "NEO human cDNA sequencing project."  
 RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

RN [3]  
 RP SEQUENCE OF 395-576 FROM N.A.  
 RC TISSUE=Brain cortex;  
 RX MEDLINE=97369492; PubMed=9225980;  
 RA Wargolis R.L., Abraham W.R., Gatchell S.B., Li S.-H., Kidwai A.S.,  
 RA Breschler T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;  
 RT "CDNAs with long CAG trinucleotide repeats from human brain."  
 RL Hum. Genet. 100:114-122(1997).

CC -!- FUNCTION: Transcription factor that binds the consensus DNA  
 CC sequence [GC]AAAAA. Seems to bind and regulate the promoters of  
 CC WML1, MMP3, MMP7 and COL1A1 (By similarity).  
 CC -!- SUBUNIT: Interacts with Cas (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=2;  
 CC Comment=Additional isoforms seem to exist;

CC Name=1;  
 CC IsoId=Q8TF68-1; Sequence=Displayed;

CC Name=2;  
 CC IsoId=Q8TF68-2; Sequence=VSP\_006920;

CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
 CC FINGER PROTEINS.

CC -!- SIMILARITY: Contains 8 C2H2-type zinc fingers.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; AB070238; BAB85425.1; -  
 DR EMBL; AK095734; BAC04618.1; -  
 DR EMBL; U80738; AAB91437.1; -

DR Genew; HGNC:11955; ZNF384.  
DR InterPro; IPR007087; Znf C2H2.  
DR Pfam; PF00096; zf-C2H2; 8.  
DR ProDom; PD000003; Znf C2H2; 4.  
DR SMART; SM00355; Znf C2H2; 8.  
DR PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
DR PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 8.  
KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;  
KW DNA-binding; Repeat; Alternative splicing.  
FT ZN FING 228 250  
FT ZN FING 256 278  
FT ZN FING 284 306  
FT ZN FING 317 339  
FT ZN FING 345 367  
FT ZN FING 373 397  
FT ZN FING 403 425  
FT ZN FING 433 455  
FT ZN FING 461 521  
FT DOMAIN 461 521  
FT DOMAIN 466 499  
FT VARSPPLIC 300 360  
FT Missing (in isoform 2).  
FT /FTid=VSP\_006920.  
FT SEQUENCE 576 AA; 63091 MW; 2A152786C3C46D90 CRC64;  
Query Match 91.4%; Score 32; DB 1; Length 576;  
Best Local Similarity 83.3%; Pred. No. 99;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RPPRGR 6  
Db 189 KPPRGR 194  
RESULT 7  
Z384\_RAT  
ID Z384\_RAT STANDARD; PRT; 579 AA.  
AC Q9EQJ4; Q9EQJ2; Q9EQJ3; Q9JMJ5;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Zinc finger protein 384 (Nuclear matrix transcription factor 4)  
DE [Cas-associated zinc finger protein].  
GN ZNF384 OR NMP4 OR CIZ.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OX NCBI\_Taxid=10116;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1), SUBCELLULAR LOCATION, DNA-BINDING, AND  
RP INTERACTION WITH CAS.  
RX MEDLINE=20136045; PubMed=10669742;  
RA Nakamoto T., Yamagata T., Sakai R., Ogawa S., Honda H., Ueno H.,  
RA Hirano N., Yazaki Y., Hirai H.;  
RT "CIZ, a zinc finger protein that interacts with p130cas and activates  
RT the expression of matrix metalloproteinases";  
RT Mol. Cell. Biol. 20:1649-1658 (2000).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).  
RX STRAIN=Sprague-Dawley;  
RX MEDLINE=21024193; PubMed=11149472;  
RA Thunyakitpisal P., Alvarez M., Iokunaga K., Onyia J.E., Hock J.,  
RA Ohashi N., Feister H., Rhodes S.J., Bigwell J.P.;  
RT "Cloning and functional analysis of a family of nuclear matrix  
RT transcription factors (NP/NMP4) that regulate type I collagen  
RT expression in osteoblasts";  
RL J. Bone Miner. Res. 16:10-23 (2001).  
CC -!- FUNCTION: Transcription factor that binds the consensus DNA  
CC sequence [GC]AATAA. Seems to bind and regulate the promoters of  
CC MEF1, MZF3, MZF7 and COL1A1.  
CC -!- SUBUNIT: Interacts with Cas.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=3;  
CC Comment=Additional isoforms seem to exist;

CC Name=1;  
CC IsoId=Q9EQJ4-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q9EQJ4-2; Sequence=VSP\_006921;  
CC Name=3;  
CC IsoId=Q9EQJ4-3; Sequence=VSP\_006922;  
CC -!- TISSUE SPECIFICITY: Expressed in osteocytes, osteoblasts, and  
CC chondrocytes in bone.  
CC -!- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-  
CC FINGER PROTEINS.  
CC -!- SIMILARITY: Contains 8 C2H2-type zinc fingers.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; AB019281; BA889664.1; -;  
CC EMBL; AF216804; AAC40582.1; -;  
CC EMBL; AF216805; AAC40583.1; -;  
CC EMBL; AF216806; AAC40584.1; -;  
CC HSSP; P08153; 1ZFD.  
CC TRANSFAC; T05136; -;  
CC TRANSFAC; T05137; -;  
CC TRANSFAC; T05138; -;  
CC TRANSFAC; T05141; -;  
CC TRANSFAC; T05142; -;  
CC InterPro; IPR007087; Znf\_C2H2.  
CC Pfam; PF00096; zf-C2H2; 8.  
CC ProDom; PD000003; Znf\_C2H2; 4.  
CC SMART; SM00355; Znf\_C2H2; 8.  
CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 8.  
CC PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 8.  
KW Transcription regulation; Zinc-finger; Metal-binding; Nuclear protein;  
KW DNA-binding; Repeat; Alternative splicing.  
FT ZN FING 229 251  
FT ZN FING 257 279  
FT ZN FING 285 307  
FT ZN FING 318 340  
FT ZN FING 346 368  
FT ZN FING 374 398  
FT ZN FING 404 426  
FT ZN FING 434 456  
FT DOMAIN 462 524  
FT DOMAIN 467 506  
FT VARSPPLIC 103 118  
FT Missing (in isoform 2).  
FT /FTid=VSP\_006921.  
FT Missing (in isoform 3).  
FT /FTid=VSP\_006922.  
FT CONFLICT 178 179 GG -> RS (IN REF. 1).  
FT CONFLICT 576 577 LA -> WP (IN REF. 1).  
SQ SEQUENCE 579 AA; 63139 MW; FBC242E0D1050C45 CRC64;  
Query Match 91.4%; Score 32; DB 1; Length 579;  
Best Local Similarity 83.3%; Pred. No. 99;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RPPRGR 6  
Db 190 KPPRGR 195  
RESULT 8  
CYSN MYCTU  
ID CYSN MYCTU STANDARD; PRT; 614 AA.  
AC Q10600;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE CysN/cysc bifunctional enzyme [includes: Sulfate adenylyltransferase



DE subunit 1 (EC 2.7.7.4) (Sulfate adenylyltransferase) (SAT) (ATP-  
DE sulfurylase large subunit); Adenylylsulfate kinase (EC 2.7.1.25) (APS  
DE kinase) (ATP adenosine-5'-phosphosulfate 3'-phosphotransferase)).  
GN CYSNC OR CYSN OR RV1286 OR MT1324 OR MTCV373.05.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI\_TaxID=1773;  
FN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37RV;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Ruster S., Seeger K., Skelton S., Squares S., Squares R.,  
RA Rulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RT complete genome sequence.";  
RL Nature 393:537-544(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CDC 1551 / Oshkosh;  
RX MEDLINE=22206494; PubMed=12218036;  
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,  
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,  
RA Kolonay J.F., Nelson W.C., Umayam L.A., Emdinova M., Salzberg S.L.,  
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,  
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RT laboratory strains.";  
RL J. Bacteriol. 184:5479-5490(2002).  
RC -!- FUNCTION: ATP sulfurylase may be the GTPase, regulating ATP  
CC sulfurylase activity (by similarity).  
CC -!- FUNCTION: APS kinase catalyzes the synthesis of activated sulfate  
CC (by similarity).  
CC -!- CATALYTIC ACTIVITY: ATP + sulfate = diphosphate + adenylylsulfate.  
CC -!- PHOSPHOADENYLYLSULFATE.  
CC -!- PATHWAY: Sulfate activation; cysteine biosynthesis reductive  
CC branch; first step.  
CC -!- PATHWAY: Sulfate activation; cysteine biosynthesis reductive  
CC branch; second step.  
CC -!- SUBUNIT: Heterodimer composed of cysD, the smaller subunit, and  
CC cysN (by similarity).  
CC -!- SIMILARITY: In the N-terminal section; belongs to the GTP-binding  
CC elongation factor family. CysN/nodQ subfamily.  
CC -!- SIMILARITY: In the C-terminal section; Belongs to the APS kinase  
CC family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; Z73419; CAA97752.1; -.  
CC EMBL; AF007007; AAK45585.1; -.  
CC PIR; B70772; B70772.  
CC HSP; P02990; 1ETU.  
CC TIGR; M11324; -.  
CC TubercuList; RV1286; -.  
CC HAMAP; MF 00062; fused; 1.  
CC HAMAP; MF 00065; fused; 1.  
CC InterPro; IPR002891; APS kinase.  
CC InterPro; IPR000795; EF GTPbind.  
CC InterPro; IPR004161; EFTU D2.  
CC InterPro; IPR009001; Elong\_init\_C.

DR InterPro; IPR009000; Translat\_factor.  
DR Pfam; P201583; APS kinase: 1.  
DR Pfam; P000009; GTP\_EFTU\_D2; 1.  
DR Pfam; PF03144; GTP\_EFTU\_D2; 1.  
DR PRINTS; PR00315; ELONGATNFCT.  
DR ProDom; PD002350; APS kinase; 1.  
DR TIGRFAMs; TIGR00455; apsk; 1.  
DR PROSITE; PS00301; EFACTOR\_GTP; 1.  
KW Cysteine biosynthesis; Transferase; Nucleotidyltransferase;  
KW GTP-binding; Kinase; ATP-binding; Multifunctional enzyme;  
KW Complete proteome. 441  
FT DOMAIN 1 441 SULFATE ADENYLYL TRANSFERASE.  
FT FT 442 614 ADENYLYLSULFATE KINASE.  
FT NP\_BIND 11 18 GTP (BY SIMILARITY).  
FT NP\_BIND 88 92 GTP (BY SIMILARITY).  
FT NP\_BIND 143 146 GTP (BY SIMILARITY).  
FT NP\_BIND 450 457 ATP (POTENTIAL).  
FT ACT\_SITE 524 524 PHOSPHOSERINE INTERMEDIATE  
FT SEQUENCE 614 AA; 67839 MW; 2C3709C8B91867C4 CRC64;  
SQ  
Query Match 91.4%; Score 32; DB 1; Length 614;  
Best Local Similarity 83.3%; Pred. No. 1.1e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 RPPRGR 6  
Db 439 RPPRGR 444  
RESULT 9  
NOSI\_MOUSE STANDARD; PRT; 1429 AA.  
ID NOSI\_MOUSE Q920J4; Q64208;  
AC 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Nitric-oxide synthase, brain (EC 1.14.13.39) (NOS, type I) (Neuronal  
DE NOS) (N-NOS) (nNOS) (Constitutive NOS) (NC-NOS) (bNOS).  
GN NOS1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS N-NOS-1 AND N-NOS-2).  
RC STRAIN=BALB/c; Tissue=Brain;  
RX MEDLINE=9331228; PubMed=7686743;  
RA Ogura T., Yokoyama T., Fujisawa H., Kurashima Y., Esumi H.;  
RT "Structural diversity of neuronal nitric oxide synthase mRNA in the nervous  
RT system.";  
RL Biochem. Biophys. Res. Commun. 193:1014-1022(1993).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM NNOS MU).  
RC TISSUE=Skeletal muscle;  
RX MEDLINE=96212184; PubMed=8628668;  
RA Silvagno F., Xia H., Bredt D.S.;  
RT "Neuronal nitric-oxide synthase-mu, an alternatively spliced isoform  
RT expressed in differentiated skeletal muscle.";  
RL J. Biol. Chem. 271:11204-11208(1996).  
RN [3]  
RP ALTERNATIVE SPLICING (ISOFORMS NNOS BETA; NNOS GAMMA AND NNOS MU).  
RX MEDLINE=97351924; PubMed=9208206;  
RA Brenman J.E., Xia H., Chao D.S., Black S.M., Bredt D.S.;  
RT "Regulation of neuronal nitric oxide synthase through alternative  
RT transcripts.";  
RL Dev. Neurosci. 19:224-231(1997).  
RN [4]  
RP INTERACTION WITH DLG4.  
RX MEDLINE=20090929; PubMed=10623522;  
RA Tochio H., Hung F., Li M., Bredt D.S., Zhang M.;  
RT "Solution structure and backbone dynamics of the second PDZ domain of  
RT postsynaptic density-95.";





```
FT METAL 420 420 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
FT DOMAIN 730 750 CALMODULIN-BINDING (POTENTIAL).
FT NP_BIND 886 917 FVN (PYRIMIDINE PART) (BY SIMILARITY).
FT NP_BIND 1032 1043 FAD (ADP PART) (BY SIMILARITY).
FT NP_BIND 1175 1185 FAD (FLAVIN PART) (BY SIMILARITY).
FT NP_BIND 1250 1268 NADP (RIBOSE PART) (BY SIMILARITY).
FT NP_BIND 1348 1363 NADP (ADP PART) (BY SIMILARITY).
FT VARSPLIC 1 336 Missing (in isoform 3).
FT VARSPLIC 285 407 /FtId=VSP 003571.
FT TLTGCTEYICGMSIMHPSCHARRDEDRTKQLFPLAKEF
IDQYSSIKRFGSKAHLERLEVNKEIDTTTQJUKOTELI
-> MKRLITEGFGVQGRSHNPPSPPOMAAPSVH
ASSRGTGLRWFSTPTSTRAHWRKDALSTSAWAPSCILL
SMOGLKTAQKDSLSPLKLLINTIHLKDLAPKPTWKG
WKR (in isoform 4).
FT VARSPLIC 408 1434 /FtId=VSP 003572.
Missing (in isoform 4).
FT VARSPLIC 509 613 /FtId=VSP 003573.
Missing (in isoform 2).
FT CONFLICT 131 131 K -> E (IN REF. 4).
FT CONFLICT 178 184 LAPRPG -> WQAPR (IN REF. 3 AND 4).
FT CONFLICT 492 493 QP -> HR (IN REF. 3).
FT CONFLICT 549 549 V -> L (IN REF. 3).
FT CONFLICT 563 563 G -> A (IN REF. 3).
FT CONFLICT 1407 1407 Y -> I (IN REF. 3).
SQ SEQUENCE 1434 AA; 160970 MW; 99235793B953BF37 CRC64;

Query Match 91.4%; Score 32; DB 1; Length 1434;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPPRGR 6
Db 516 KPPRGR 521

RESULT 11
ID DPM1 MOUSE STANDARD; PRT; 260 AA.
AC Q70152; Q9D829;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE Dolichol-phosphate mannosyltransferase (EC 2.4.1.83) (Dolichol-
DE phosphate mannosyltransferase) (Dolichyl-phosphate beta-D-
DE mannosyltransferase) (Mannose-P-dolichol synthase) (MPD synthase) (DPM
DE synthase).
GN DPM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98204925; PubMed=9535917;
RA Tomita S., Inoue N., Maeda Y., Ohishi K., Takeda J., Kinoshita T.;
RT "A homologue of Saccharomyces cerevisiae Dpm1p is not sufficient for
RT synthesis of dolichol-phosphate-mannose in mammalian cells.";
RL J. Biol. Chem. 273:9249-9254(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, Lung, and Small intestine;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Komno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saico T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Cissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nakado I., Pesole G., Quackenbush J.,
RA Schriml L.N., Staebli F., Suzuki R., Tomita M., Wagner L., Washio T.,
```

```
RA Sakai K., Okido T., Furuno M., Aono H., Baidarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
RA Brownstein M.J., Bult C., Gletcher C., Fujita M., Gariboldi M.F.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Havaahizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -I- FUNCTION: Transfers mannose from GDP-mannose to dolichol
CC monophosphate to form dolichol phosphate mannose (Dol-P-Man) which
CC is the mannosyl donor in pathways leading to N-glycosylation,
CC glycosyl phosphatidylinositol membrane anchoring, and O-
CC mannosylation of proteins.
CC -I- CATALYTIC ACTIVITY: GDP-mannose + dolichyl phosphate = GDP +
CC dolichyl D-mannosyl phosphate.
CC -I- PATHWAY: Glycosylation.
CC -I- SUBCELLULAR LOCATION: Endoplasmic reticulum.
CC -I- SIMILARITY: Belongs to the glycosyltransferase family 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB004789; BAA25759.1; -
CC EMBL; AK005268; BAA23920.1; -
CC EMBL; AK004834; BAA23602.1; -
CC EMBL; AK008548; BAA25735.1; -
CC MGD; MGI:1330239; Dpm1.
CC GO; GO:0000030; F:mannosyltransferase activity; IMP.
CC InterPro; IPR001173; Glyco_transf_2; 1.
CC Pfam; PF00535; Glycosyltransferase; Endoplasmic reticulum.
KW Transferase; Glycosyltransferase; E -> G (IN REF. 2; BAA25735).
FT CONFLICT 76 76 E -> G (IN REF. 2; BAA25735).
FT CONFLICT 122 122 S -> Y (IN REF. 2; BAA25735).
SQ SEQUENCE 260 AA; 29174 MW; 8CB1BCDF1977EBA6 CRC64;

Query Match 88.6%; Score 31; DB 1; Length 260;
Best Local Similarity 83.3%; Pred. No. 67;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RPPRGR 6
Db 15 RPPQGR 20

RESULT 12
ID DPM1 CRIGR STANDARD; PRT; 266 AA.
AC Q9WU83;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Dolichol-phosphate mannosyltransferase (EC 2.4.1.83) (Dolichol-
DE phosphate mannosyltransferase) (Dolichyl-phosphate beta-D-
DE mannosyltransferase) (Mannose-P-dolichol synthase) (MPD synthase) (DPM
DE synthase).
GN DPM1.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RA Pu L., Sococca J.R., Walker B.K., Wu J.S., Krag S.S.;
```

RT "Mutation in B4-2-1 CHO cells defective in MPD synthase activity.";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Transfers mannose from GDP-mannose to dolichol  
 CC monophosphate to form dolichol phosphate mannose (Dol-P-Man) which  
 CC is the mannoseyl donor in pathways leading to N-glycosylation,  
 CC glycosyl phosphatidylinositol membrane anchoring, and O-  
 CC mannosylation of proteins (By similarity).  
 CC -!- CATALYTIC ACTIVITY: GDP-mannose + dolichyl phosphate = GDP +  
 CC dolichyl D-mannosyl phosphate.  
 CC -!- PATHWAY: Glycosylation.  
 CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum (By similarity).  
 CC -!- SIMILARITY: Belongs to the glycosyltransferase family 2.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; AF121895; AAD30975.1; -;  
 CC InterPro; IPR001173; Glyco\_transf\_2.  
 CC Pfam; PF00535; Glycosyltransferase; Endoplasmic reticulum.  
 CC KW Transferase; Glycosyltransferase; Endoplasmic reticulum.  
 CC SQ SEQUENCE 266 AA; 29654 MW; 4AFB37EA3AC3329D CRC64;  
 CC  
 CC Query Match 88.6%; Score 31; DB 1; Length 266;  
 CC Best Local Similarity 83.3%; Pred. No. 69;  
 CC Matches 5; Conservative 1; Mismatches 0; Gaps 0;  
 CC  
 CC QY 1 RPPQGR 6  
 CC |||||  
 CC Db 21 RPPQGR 26  
 CC  
 CC RESULT 13  
 CC ROC2 ARATH  
 CC ID ROC2 ARATH STANDARD; PRT; 342 AA.  
 CC AC Q43349; Q94BT1; Q9LFH2;  
 CC DT 16-OCT-2001 (Rel. 40, Created)  
 CC DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 CC DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 CC DE 29 kDa ribonucleoprotein, chloroplast precursor (RNA-binding protein  
 CC cp29).  
 CC GN RBP29 OR AT3G53460 OR P4P12.160.  
 CC OS Arabidopsis thaliana (Mouse-ear cress).  
 CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 CC OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 CC NCBI\_Taxid=3702;  
 CC [1]  
 CC RN SEQUENCE FROM N.A. (ISOFORM 2).  
 CC RC STRAIN=cv. Columbia; TISSUE=leaf;  
 CC EX MEDLINE=95201246; Pubmed=7894017;  
 CC RA Chca M., Sugita M., Sugiura M.;  
 CC RT "Three types of nuclear genes encoding chloroplast RNA-binding  
 CC proteins (cp29, cp31 and cp33) are present in Arabidopsis thaliana:  
 CC RT presence of cp31 in chloroplasts and its homologue in  
 CC RT nuclei/cytoplasms.";  
 CC RL Plant Mol. Biol. 27:529-539 (1995).  
 CC [2]  
 CC RN SEQUENCE FROM N.A.  
 CC RC STRAIN=cv. Columbia;  
 CC RX MEDLINE=21016720; Pubmed=11130713;  
 CC RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,  
 CC RA Fatmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,  
 CC RA Delseray M., Bourry M., Grivell L.A., Mache R., Puigdomenech P.,  
 CC RA De Simone V., Choigne N., Artiguenave F., Robert C., Brottier P.,  
 CC RA Wincker P., Cattolico L., Weissbach J., Saurin W., Quetier F.,  
 CC RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,  
 CC RA Wurmbach E., Drzonek H., Erfle H., Jordan N., Bangert S.,  
 CC RA Wiedelmann R., Kranz H., Voss H., Holland R., Brandt P., Nyakatura G.,

Vezi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,  
 Conrad A., Hornischer K., Kauer G., Leebert T.-H., Nordsiek G.,  
 Reichelt J., Schaffe M., Schoen O., Barques M., Terol J., Climent J.,  
 Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,  
 Cooke R., Laude M., Berger-Liauro C., Purnelle B., Masuy D.,  
 de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,  
 Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,  
 Mannhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Meves H.-W.,  
 Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,  
 Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,  
 Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,  
 Pai G., Miltischer J., Sellers P., Gill J.E., Feldblyum T.V.,  
 Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,  
 Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,  
 Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,  
 Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,  
 Watanabe A., Yamada M., Yasuda M., Tabata S.;  
 "Sequence and analysis of chromosome 3 of the plant Arabidopsis  
 thaliana.";  
 RT Nature 408:820-822 (2000).  
 RL [3]  
 RN SEQUENCE FROM N.A. (ISOFORM 1).  
 RP STRAIN=cv. Columbia;  
 RX MEDLINE=22954850; Pubmed=14593172;  
 RA Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,  
 Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,  
 Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,  
 Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,  
 Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,  
 Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,  
 Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,  
 Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,  
 Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,  
 Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,  
 Satou M., Tamse R., Vaysberg M., Wallenberg E.K., Wong C., Yamamura Y.,  
 Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
 "Empirical analysis of transcriptional activity in the Arabidopsis  
 genome.";  
 RT Science 302:842-846 (2003).  
 RL [4]  
 RN SEQUENCE FROM N.A. (ISOFORM 1).  
 RP Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,  
 Feldmann K.A.;  
 RT "Full-length cDNA from Arabidopsis thaliana.";  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: Could be involved in splicing and/or processing of  
 CC chloroplast RNA's.  
 CC -!- SUBCELLULAR LOCATION: Chloroplast.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1;  
 CC IsoId=Q43349-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=Q43349-2; Sequence=VSP\_009110;  
 CC -!- SIMILARITY: Contains 2 RNA recognition motif (RRM) domains.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; D31710; BAA06518.1; -;  
 CC ENBL; D31711; BAA06519.1; -;  
 CC ENBL; ALI32966; CAB67653.1; -;  
 CC ENBL; AY039909; AAK64013.1; -;  
 CC ENBL; AY077674; AAL76152.1; -;  
 CC ENBL; AY087840; AAM65393.1; -;  
 CC PIR; S53490; S53490.  
 CC HSP; P09651; LHAI.

```

DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 2.
DR SMART: SMO0360; RRM; 2.
DR PROSITE: PS00300; RRM_RNP_1; 2.
DR mRNA processing: Ribonucleoprotein; Chloroplast; RNA-binding; Repeat;
Transit peptide; Alternative splicing.
FT TRANSIT 1 65 CHLOROPLAST (POTENTIAL).
FT CHAIN 66 342 29 KDA RIBONUCLEOPROTEIN.
FT DOMAIN 99 177 RNA-BINDING (RRM) 1.
FT DOMAIN 257 335 RNA-BINDING (RRM) 2.
FT DOMAIN 178 256 LINKER (GLY-RICH).
FT VARSPLIC 220 227 Missing (in isoform 2).
FT /FTIG=VSP 009110.
SQ SEQUENCE 342 AA; 36007 MW; CAC3410E72410988 CRC64;

Query Match 88.6%; Score 31; DB 1; Length 342;
Best Local Similarity 83.3%; Pred. No. 89;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6
|||:|
Db 336 RPPRQG 341

RESULT 14
SNA4_STRPR STANDARD; PRT; 421 AA.
ID _SNA4_STRPR STANDARD; PRT; 421 AA.
AC P54951;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Pristinamycin IIA synthase subunit A (PIIA synthase subunit A).
GN SNA4.
OS Streptomyces pristinaespiralis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=38300;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SP92;
RX MEDLINE=95394837; PubMed=7665509;
RA Blanc V., Lagneaux D., Didier P., Gil P., Lacroix P., Crouzet J.;
RT "Cloning and analysis of structural genes from Streptomyces
RT pristinaespiralis encoding enzymes involved in the conversion of
RT pristinamycin IIB to pristinamycin IIA (PIIA): PIIA synthase and
RT NADH:riboflavin 5'-phosphate oxidoreductase.";
RL J. Bacteriol. 177:5206-5214(1995).
RN [2]
RP SEQUENCE OF 1-18 AND 364-383.
RX MEDLINE=95394836; PubMed=7665508;
RA Thibaut D., Ratet N., Bisch D., Faucher D., Debussche L., Blanche F.;
RT "Purification of the two-enzyme system catalyzing the oxidation of
RT the D-proline residue of pristinamycin IIB during the last step of
RT pristinamycin IIA biosynthesis.";
RL J. Bacteriol. 177:5198-5205(1995)
CC -!- FUNCTION: CATALYZES THE OXIDATION OF THE PROLINE RESIDUE OF
CC PRISTINAMYCIN IIB (PIIB) TO PRISTINAMYCIN IIA (PIIA).
CC -!- COFACTOR: FMN.
CC -!- SUBUNIT: HETERODIMER OF TWO SUBUNITS, SNA4 AND SNA5.
CC -!- SIMILARITY: BELONGS TO THE NTRA/SNA4/SOXA(DSZA) FAMILY OF
CC MONOOXYGENASES.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; U21215; AAA83563.1; -.
CC InterPro: IPR002103; Bac_luciferase.

```

---

```

DR Pfam: PF00296; bac_luciferase; 1.
DR Oxidoreductase; Monooxygenase; Flavoprotein; FMN.
FT INIT MET 0
SQ SEQUENCE 421 AA; 46373 MW; 341AD008A939CEFA CRC64;

Query Match 88.6%; Score 31; DB 1; Length 421;
Best Local Similarity 83.3%; Pred. No. 11e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RPPRGR 6
|||:|
Db 192 RPPQGR 197

RESULT 15
TMC2_MOUSE STANDARD; PRT; 888 AA.
ID TMC2_MOUSE STANDARD; PRT; 888 AA.
AC Q8R4P4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transmembrane cochlear-expressed protein 2.
GN TMC2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, inner ear, and Testis;
RX MEDLINE=21918583; PubMed=11850618;
RA Kurima K., Peters L.M., Yang Y., Riazuddin S., Ahmed Z.M., Naz S.,
RA Arshad D., Drury S., Mo J., Makishima T., Ghosh M., Menon P.S.N.,
RA Deshmukh D., Oddoux C., Ostrer H., Khan S., Raizuddin S.,
RA Deininger P.L., Hampton L.L., Sullivan S.L., Battey J.F.,
RA Keats B.J.B., Wilcox E.R., Friedman T.B., Griffith A.J.;
RT "Dominant and recessive deafness caused by mutations of a novel gene,
RT TMC1, required for cochlear hair-cell function.";
RL Nat. Genet. 30:277-284(2002).
CC -!- FUNCTION: May be required for the normal function of cochlear hair
CC cells.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- TISSUE SPECIFICITY: Inner ear.
CC -!- DEVELOPMENTAL STAGE: Expressed at low, constant levels in temporal
CC bone from embryonic day 14 to day 1 after birth. Increases by 8 to
CC 16-fold at day 5, 10 and 20.
CC -!- SIMILARITY: Belongs to the TMC family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AF417581; AAL86402.1; -.
CC MGd; MG1:2151017; Tmc2.
DR Transmembrane.
KW DOMAIN 1 245 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 246 266 POTENTIAL.
FT DOMAIN 267 318 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 319 339 POTENTIAL.
FT DOMAIN 340 412 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 413 433 POTENTIAL.
FT DOMAIN 434 490 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 491 511 POTENTIAL.
FT DOMAIN 512 575 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 576 596 POTENTIAL.
FT DOMAIN 597 733 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 733 753 POTENTIAL.
FT DOMAIN 754 888 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 888 932 GLU/ASP/LYS/ARG-RICH (HIGHLY CHARGED).

```

SQ SEQUENCE 888 AA; 101134 MW; 6D834D7987768FA7 CRC64;  
Query Match 88.6%; Score 31; DB 1; Length 888;  
Best Local Similarity 83.3%; Pred. No. 2.3e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RPPRGR 6  
Db 857 QPPRGR 862  
Search completed: April 6, 2004, 16:08:01  
Job time : 4.86916 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:52:34 ; Search time 5.60748 Seconds  
(without alignments)  
85.771 Million cell updates/sec

Title: US-10-009-709-12

Perfect score: 30

Sequence: 1 PPRGR 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	36	2 C58788	bone morphogenetic
2	30	100.0	55	2 S29770	DNA-binding protein
3	30	100.0	79	2 C37914	homeotic protein C
4	30	100.0	100	2 E43550	hypothetical prote
5	30	100.0	108	1 B3AG55	virB3 protein - Ag
6	30	100.0	143	2 E72899	hypothetical prote
7	30	100.0	150	2 D72870	hypothetical prote
8	30	100.0	166	1 PIHUSC	salivary proline-r
9	30	100.0	166	2 B25372	hypothetical prote
10	30	100.0	167	2 AE2630	hypothetical prote
11	30	100.0	168	2 T46341	hypothetical prote
12	30	100.0	171	2 A27307	proline-rich phosph
13	30	100.0	178	2 A75036	hypothetical prote
14	30	100.0	188	2 S72709	Lept170 C3_229 pr
15	30	100.0	199	2 S41316	coat protein - cuc
16	30	100.0	200	2 S55609	hypothetical prote
17	30	100.0	204	2 T50572	hypothetical prote
18	30	100.0	239	2 A70626	hypothetical prote
19	30	100.0	244	2 B87556	hypothetical prote
20	30	100.0	246	2 T49305	hypothetical prote
21	30	100.0	272	2 A44110	hypothetical prote
22	30	100.0	285	2 T34836	probable dehydrata
23	30	100.0	302	2 H96811	protein F3f9.20 [i
24	30	100.0	308	2 B38234	oxoglutarate dehyd
25	30	100.0	310	2 T36542	hypothetical prote
26	30	100.0	322	2 G75291	conserved hypothet
27	30	100.0	335	2 S47304	gene M protein - r
28	30	100.0	347	2 T08954	hypothetical prote
29	30	100.0	350	1 VGBE63	glycoprotein gp63

#### ALIGNMENTS

##### RESULT 1

C58788

bone morphogenetic protein 1, BMP-1 splice form - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 09-Apr-1998 #sequence\_revision 28-Mar-1998 #text\_change 20-Apr-2000

C:Accession: C58788

R:Takahara, K.; Lyons, G.E.; Greenspan, D.S.

J. Biol. Chem. 269, 32572-32578, 1994

A>Title: Bone morphogenetic protein-1 and a mammalian tolloid homologue (mTld) are enc

A:Reference number: A58788; MUID:95096114; PMID:7798260

A:Accession: C58788

A:Molecule type: mRNA

A:Residues: 1-36 <TAK>

A:Cross-references: GB:L35280; NID:g623162; PIDN:AAC42043.1; PID:g623163

A:Experimental source: 17.5 day embryo

C:Genetics:

A:Gene: BMP-1

C:Superfamily: procollagen C-endopeptidase; astacin homology; C1r/C1s repeat homology;

C:Keywords: alternative splicing

Query Match 100.0%; Score 30; DB 2; Length 36;  
Best Local Similarity 100.0%; Pred. No. 17; Mismatches 0; Indels 0; Gaps 0;  
Matches 5; Conservative 0;

Qy 1 PPRGR 5

Db 16 PPRGR 20

##### RESULT 2

S29770

DNA-binding protein 100K chain - human (fragments)

C:Species: Homo sapiens (man)

C>Date: 25-Feb-1994 #sequence\_revision 24-May-1996 #text\_change 07-May-1999

C:Accession: S29770

R:Zhang, W.W.; Zhang, L.X.; Busch, R.K.; Farres, J.; Busch, H.

Biochem. J. 290, 267-272, 1993

A>Title: Purification and characterization of a DNA-binding heterodimer of 52 and 100

A:Reference number: S29769; MUID:93176127; PMID:8439294

A:Accession: S29770

A:Molecule type: protein

A:Residues: 1-34;35-55 <ZHA>

C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprot

C:Keywords: DNA binding; heterodimer

Query Match 100.0%; Score 30; DB 2; Length 55;  
Best Local Similarity 100.0%; Pred. No. 25; Mismatches 0; Indels 0; Gaps 0;  
Matches 5; Conservative 0;

Qy 1 PPRGR 5

Db 29 PPRGR 33



## RESULT 3

C37914  
 A:Title: homeotic protein Chox-4f - chicken (fragment)  
 C:Species: Gallus gallus (chicken)  
 C>Date: 26-Jul-1991 #sequence\_revision 26-Jul-1991 #text\_change 24-Sep-1999  
 C:Accession: C37914  
 R:Notho, T.; Noji, S.; Koyama, E.; Ohyama, K.; Myokai, F.; Kuroiwa, A.; Saito, T.; Taniguchi, T.  
 A:Title: Involvement of the Chox-4 chicken homeobox genes in determination of anteroposterior axis in Drosophila  
 A:Reference number: A37914; PMID:1672266  
 A:Accession: C37914  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-79 <NOH>  
 A:Cross-references: GB:D10289; NID:G222806; PIDN:BAA01134.1; PID:d1001604; PID:G222807;  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F:12-68/Domain: homeobox homology <HOX>

Query Match 100.0%; Score 30; DB 2; Length 79;

Best Local Similarity 100.0%; Pred. No. 36;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5

Db 6 PPRGR 10

## RESULT 4

E43550  
 A:Title: hypothetical protein E4 - human papillomavirus type 41  
 C:Species: human papillomavirus type 41  
 A:Note: host Homo sapiens (man)  
 C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 18-Nov-1994  
 C:Accession: E43550  
 R:Hirt, L.; Hirsch-Behn, A.; De Villiers, E.M.  
 A:Title: Nucleotide sequence of human papillomavirus (HPV) type 41: an unusual HPV type  
 A:Reference number: A43550  
 A:Accession: E43550  
 A:Status: translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-100 <HIR>  
 A:Cross-references: EMBL:X56147

Query Match 100.0%; Score 30; DB 2; Length 100;

Best Local Similarity 100.0%; Pred. No. 44;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5

Db 5 PPRGR 9

## RESULT 5

B3AG55  
 A:Title: virB3 protein - Agrobacterium tumefaciens plasmids pTi15955 and pTiA6  
 C:Species: Agrobacterium tumefaciens  
 C>Date: 30-Jun-1991 #sequence\_revision 30-Jun-1991 #text\_change 16-Jul-1999  
 C:Accession: S00779; C28621; C27127  
 R:Thompson, D.V.; Melchers, L.S.; Idler, K.B.; Schilperoort, R.A.; Hooykaas, P.J.J.  
 A:Title: Analysis of the complete nucleotide sequence of the Agrobacterium tumefaciens virB3 protein  
 A:Reference number: S00779; PMID:8824765; PMID:2837739  
 A:Accession: S00779  
 A:Molecule type: DNA  
 A:Residues: 1-108 <THO>  
 A:Cross-references: EMBL:X06826; NID:G39195; PIDN:CAA29974.1; PID:G39202  
 A:Experimental source: strain 15955, plasmid pTi15955  
 R:Ward, J.E.; Akiyoshi, D.E.; Regier, D.; Datta, A.; Gordon, M.P.; Nester, E.W.  
 J. Biol. Chem. 263, 5804-5814, 1988

A:Title: Characterization of the virB operon from an Agrobacterium tumefaciens Ti plasmid  
 A:Reference number: A28621; PMID:88186901; PMID:13281947

A:Accession: C28621

A:Molecule type: DNA

A:Residues: 1-108 <WAR>

A:Cross-references: GB:J03216; NID:G1196971; PIDN:AAA88647.1; PID:G1196974

A:Experimental source: plasmid pTiA6

C:Genetics:

A:Genome: plasmid

C:Superfamily: tumor-inducing plasmid pTiC58 virB3 protein

Query Match 100.0%; Score 30; DB 1; Length 108;

Best Local Similarity 100.0%; Pred. No. 48;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5

Db 101 PPRGR 105

## RESULT 6

E72859  
 A:Title: hypothetical protein APE1012 - Aeropyrum pernix (strain K1)  
 C:Species: Aeropyrum pernix  
 C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 15-Sep-2003  
 C:Accession: E72859  
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;  
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix  
 A:Reference number: A72450; PMID:99310339; PMID:10382966  
 A:Accession: E72859  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-143 <KAW>  
 A:Cross-references: DDBJ:AP000060; NID:G5104188; PIDN:BAA79997.1; PID:d1043783; PID:G5104188  
 A:Experimental source: strain K1  
 C:Genetics:

A:Gene: APE1012

Query Match 100.0%; Score 30; DB 2; Length 143;

Best Local Similarity 100.0%; Pred. No. 62;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5

Db 93 PPRGR 97

## RESULT 7

D72670  
 A:Title: hypothetical protein APE0786 - Aeropyrum pernix (strain K1)  
 C:Species: Aeropyrum pernix  
 C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Aug-1999  
 C:Accession: D72670  
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.;  
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix  
 A:Reference number: A72450; PMID:99310339; PMID:10382966  
 A:Accession: D72670  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-150 <KAW>  
 A:Cross-references: DDBJ:AP000060; NID:G5104188; PIDN:BAA79764.1; PID:d1043550; PID:G5104188  
 A:Experimental source: strain K1  
 C:Genetics:

A:Gene: APE0786

Query Match 100.0%; Score 30; DB 2; Length 150;

Best Local Similarity 100.0%; Pred. No. 65;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 PPRGR 5
Db 68 PPRGR 72

RESULT 8
PIHUSC
salivary proline-rich phosphoprotein precursor PRH2 (validated) - human
N/Alternate names: salivary acidic proline-rich protein PRH2
N/Contents: peptide P-C (basic proline-rich peptide IB-8b); proline-rich phosphoprotein
C/Species: Homo sapiens (man)
C/Date: 31-Mar-1991 #sequence_revision 12-Apr-1996 #text_change 08-Dec-2000
C/Accession: A25372; A19803; B57868; A92277; A92254; A94425; A91954; S02564; S02563; JPO
R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.
J. Biol. Chem. 260, 11123-11130, 1985
A/Title: Differential RNA splicing and post-translational cleavages in the human salivary
A/Reference number: A92492; MUID:85289325; PMID:2993301
A/Accession: A25372
A/Molecule type: mRNA
A/Residues: 1-166 <MAE>
A/Cross-references: GB:K03202; NID:G190481; PIDN:AAA60183.1; PID:G190482
R/Schlesinger, D.H.; Hay, D.I.
Int. J. Pept. Protein Res. 17, 34-41, 1981
A/Title: Primary structure of the active tryptic fragments of human and monkey salivary
A/Reference number: A91757; MUID:81191179; PMID:7228490
A/Accession: A19803
A/Molecule type: protein
A/Residues: 17-46 <SCH>
R/Kim, H.S.; Maeda, N.
J. Biol. Chem. 261, 6712-6718, 1986
A/Title: Structures of two HaellI-type genes in the human salivary proline-rich protein
A/Reference number: A57868; MUID:86196106; PMID:3009472
A/Accession: B57868
A/Molecule type: DNA
A/Residues: 1-166 <KIM>
A/Cross-references: GB:M13058; NID:G190513; PIDN:AAA98808.1; PID:G190514
R/Wong, R.S.C.; Bennick, A.
J. Biol. Chem. 255, 5943-5948, 1980
A/Title: The primary structure of a salivary calcium-binding proline-rich phosphoprotein
A/Reference number: A92277; MUID:80204368; PMID:7380845
A/Contents: protein C
A/Accession: A92277
A/Molecule type: protein
A/Residues: 17-19, N, 21-166 <WON>
A/Note: the amino-terminal 46 residues are involved with inhibiting hydroxyapatite forma
R/Wong, R.S.C.; Hofmann, T.; Bennick, A.
J. Biol. Chem. 254, 4800-4808, 1979
A/Title: The complete primary structure of a proline-rich phosphoprotein from human saliv
A/Reference number: A92254; MUID:79173237; PMID:438215
A/Contents: protein A
A/Accession: A92254
A/Molecule type: protein
A/Residues: 17-19, N, 21-122 <WOC2>
R/Schlesinger, D.H.; Hay, D.I.
in Peptides: Structure and Biological Function (Proc. Sixth Amer. Peptide Symp.), Gross,
A/Title: Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent i
A/Reference number: A94425
A/Accession: A94425
A/Molecule type: protein
A/Residues: 17-122 <SC2>
A/Note: the authors call this protein PRP-4
R/Isemura, S.; Saitoh, E.; Sanada, K.
J. Biochem. 87, 1071-1077, 1980
A/Title: The amino acid sequence of a salivary proline-rich peptide, P-C, and its relati
A/Reference number: A91954; MUID:80227634; PMID:7390979
A/Contents: peptide P-C
A/Accession: A91954
A/Molecule type: protein
A/Residues: 123-166 <ISE>
R/Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Minaguchi, K.; Madapallimattam, G.; Schluck
Biochem. J. 255, 15-21, 1988
A/Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-
A/Reference number: S02562; MUID:89061650; PMID:3196309

```

A/Accession: S02564  
A/Molecule type: protein  
A/Residues: 17-166 <HAY>  
A/Accession: S02563  
A/Molecule type: protein  
A/Residues: 47-71 <HA2>  
R/Schlesinger, D.H.; Hay, D.I.  
Int. J. Pept. Protein Res. 27, 373-379, 1986  
A/Title: Complete structure of a proline-rich phosphoprotein, PRP-2, an inhib  
A/Reference number: JPO106; MUID:86222916; PMID:3710693  
A/Accession: JPO106  
A/Molecule type: protein  
A/Residues: 17-161, Q, 163-166 <SC3>  
A/Experimental source: parotid gland  
R/Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
Biochemistry 30, 3351-3356, 1991  
A/Title: Basic proline-rich proteins from human parotid saliva: relationships of the c  
A/Reference number: A93355; MUID:91190884; PMID:1849422  
A/Accession: G38355  
A/Status: preliminary  
A/Molecule type: protein  
A/Residues: 123-166 <KAU>  
R/Robinson, R.; Kaufman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
Biochem. J. 263, 497-503, 1989  
A/Title: Primary structure and possible origin of the non-glycosylated basic proline-r  
A/Reference number: S06153; MUID:90088384; PMID:2688632  
A/Accession: S06153  
A/Molecule type: protein  
A/Residues: 123-166 <ROB>  
C/Comment: The proposed biological functions are a highly potent inhibitor of crystal  
C/Genetics:  
A/Gene: GDB:PRH2  
A/Cross-references: GDB:119516; OMIM:168790  
A/Map position: 12p13.2-12p13.2  
A/Introns: 22/1; 34/1  
C/Superfamily: proline-rich protein  
C/Keywords: calcium binding; phosphoprotein; pyroglutamic acid; saliva  
F/1-16/Domain: signal sequence #status predicted <SG>  
F/17-166/Product: protein C #status experimental <PRC>  
F/17-122/Product: protein A #status experimental <PRA>  
F/17-46/Region: apatitic mineral binding  
F/47-71/Product: PRP-3 #status experimental <PRP3>  
F/123-166/Product: peptide P-C #status experimental <PPC>  
F/17/Modified site: pyroglutamic acid (Gln) (in mature form) #status experim  
F/24,38/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 100.0%; Score 30; DB 1; Length 166;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Oy 1 PPRGR 5  
Db 120 PPRGR 124

RESULT 9  
B25372  
salivary proline-rich phosphoprotein precursor PRH1 (allele PIF) - human  
C/Species: Homo sapiens (man)  
C/Date: 29-Aug-1987 #sequence\_revision 29-Aug-1987 #text\_change 20-Aug-1999  
C/Accession: B25372; A57868; S02562; G38355; S06153; B27307  
R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
J. Biol. Chem. 260, 11123-11130, 1985  
A/Title: Differential RNA splicing and post-translational cleavages in the human saliv  
A/Reference number: A92492; MUID:85289325; PMID:2993301  
A/Accession: B25372  
A/Molecule type: mRNA  
A/Residues: 1-166 <MAE>  
A/Cross-references: GB:K03203; NID:G190483; PIDN:AAA60184.1; PID:G190484  
R/Kim, H.S.; Maeda, N.  
J. Biol. Chem. 261, 6712-6718, 1986  
A/Title: Structures of two HaellI-type genes in the human salivary proline-rich protei  
A/Reference number: A57868; MUID:86196106; PMID:3009472

A:Accession: A57868  
 A:Molecule type: DNA  
 A:Residues: 1-166 <KIM>  
 A:Cross-references: GB:M13057; NID:G190511; PIDN:AAA98807.1; PID:G190512  
 R:Hay, D.I.; Bennick, A.; Schlesinger, D.H.; Managuchi, K.; Madapallimattam, G.; Schluck  
 Biochem. J. 255, 15-21, 1988  
 A:Title: The primary structures of six human salivary acidic proline-rich proteins (PRP-  
 A:Reference number: S02562; MUID:89061650; PMID:3196309  
 A:Accession: S02562  
 A:Molecule type: protein  
 A:Residues: 47-71 <HAY>  
 R:Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A:Title: Basic proline-rich proteins from human parotid saliva: relationships of the cov  
 A:Reference number: A38355; MUID:91190884; PMID:1849422  
 A:Accession: G38355  
 A:Molecule type: protein  
 A:Residues: 123-166 <KAU>  
 R:Robinson, R.; Kauffman, D.L.; Wayne, M.M.Y.; Blum, M.; Bennick, A.; Keller, P.J.  
 Biochem. J. 263, 497-503, 1989  
 A:Title: Primary structure and possible origin of the non-glycosylated basic proline-rich  
 A:Reference number: S06153; MUID:90088384; PMID:2688632  
 A:Accession: S06153  
 A:Molecule type: protein  
 A:Residues: 123-166 <ROB>  
 R:Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A:Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich pro  
 A:Reference number: A27307; MUID:88074309; PMID:3687941  
 A:Contents: allele Pa  
 A:Accession: B27307  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 17-41, 'L', 43-118, 'C', 120-166 <AZE>  
 A:Cross-references: EMBL:K03203  
 C:Genetics:  
 A:Gene: GDB:PRH1  
 A:Cross-references: GDB:119515; OMIM:168730  
 A:Map position: 12p13.2-12p13.2  
 A:Introns: 22/1, 34/1  
 C:Superfamily: proline-rich protein  
 C:Keywords: phosphoprotein; saliva; tandem repeat

Query Match 100.0%; Score 30; DB 2; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
 |||||  
 Db 120 PPRGR 124

RESULT 10  
 AE2630  
 hypothetical protein Atu0440 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
 C:Species: Agrobacterium tumefaciens  
 C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 18-Nov-2002  
 C:Accession: AE2630  
 R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.R.; Chen, Y.; Woo, I  
 erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell  
 ; Karp, P.; Romero, P.; Zhang, S.  
 Science 294, 2317-2323, 2001  
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,  
 star, E.W.  
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
 A:Reference number: AB2577; MUID:21608550; PMID:11743193  
 A:Accession: AE2630  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-167 <KUR>  
 A:Cross-references: GB:AE008688; PIDN:AAJ41459.1; PID:G17738783; GSPDB:GN00186  
 A:Experimental source: strain C58 (Dupont)  
 C:Genetics:

A:Gene: Atu0440  
 A:Map position: circular chromosome

Query Match 100.0%; Score 30; DB 2; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
 |||||  
 Db 69 PPRGR 73

RESULT 11  
 T46341  
 hypothetical protein DKFZp434B0914.1 - human  
 C:Species: Homo sapiens (man)  
 C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 04-Feb-2000  
 C:Accession: T46341  
 R:Kocher, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
 submitted to the Protein Sequence Database, January 2000  
 A:Reference number: Z23037  
 A:Accession: T46341  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-168 <AAA>  
 A:Cross-references: EMBL:AL137653  
 A:Experimental source: adult testis; clone DKFZp434B0914  
 C:Genetics:  
 A:Note: DKFZp434B0914.1

Query Match 100.0%; Score 30; DB 2; Length 168;  
 Best Local Similarity 100.0%; Pred. No. 72;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
 |||||  
 Db 56 PPRGR 60

RESULT 12  
 A27307  
 proline-rich phosphoprotein (gene PRH1, Db allele) - human  
 N:Alternate names: salivary acidic proline-rich protein  
 C:Species: Homo sapiens (man)  
 C:Date: 30-Jun-1988 #sequence\_revision 30-Jun-1988 #text\_change 29-Aug-1997  
 C:Accession: A27307  
 R:Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A:Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich p  
 A:Reference number: A27307; MUID:88074309; PMID:3687941  
 A:Accession: A27307  
 A:Status: nucleic acid sequence not shown  
 A:Molecule type: DNA  
 A:Residues: 1-171 <AZE>  
 A:Cross-references: EMBL:K03203  
 C:Genetics:  
 A:Gene: GDB:PRH1  
 A:Cross-references: GDB:119515; OMIM:168730  
 A:Map position: 12p13.2-12p13.2  
 C:Superfamily: proline-rich protein  
 C:Keywords: phosphoprotein

Query Match 100.0%; Score 30; DB 2; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 73;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
 |||||  
 Db 125 PPRGR 129

RESULT 13  
 A75036

hypothetical protein PAB1533 - Pyrococcus abyssi (strain Orsay)  
 C/Species: Pyrococcus abyssi  
 C/Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 20-Jun-2000  
 C/Accession: A75036  
 R:anonymous, Genoscope  
 submitted to the EMBL Data Library, July 1999  
 A/Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure  
 A/Reference number: A75001  
 A/Accession: A75036  
 A/Molecule type: DNA  
 A/Status: preliminary  
 A/Residues: 1-178 <RAW>  
 A/Cross-references: GB:AJ248287; GB:AL056836; NID:G5458657; PIDN:CA850182.1; PID:G545869  
 A/Experimental source: strain Orsay  
 C/Genetics:  
 A/Gene: PAB1533

Query Match 100.0%; Score 30; DB 2; Length 178;  
 Best Local Similarity 100.0%; Pred. No. 76;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
 |||||  
 Db 28 PPRGR 32

## RESULT 14

S72709  
 Lepb1170\_C3\_229 protein - Mycobacterium leprae  
 C/Species: Mycobacterium leprae  
 C/Date: 19-Mar-1997 #sequence\_revision 25-Apr-1997 #text\_change 22-Oct-1999  
 C/Accession: S72709  
 R:Smith, D.R.; Robison, K.  
 submitted to the EMBL Data Library, November 1993  
 A/Description: Mycobacterium leprae cosmid B1170.  
 A/Reference number: S72693  
 A/Accession: S72709  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-198 <SMI>  
 A/Cross-references: EMBL:U00010; NID:G466780; PIDN:AA17073.1; PID:G466797  
 C/Genetics:  
 A/Start codon: GTG

Query Match 100.0%; Score 30; DB 2; Length 198;  
 Best Local Similarity 100.0%; Pred. No. 84;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
 |||||  
 Db 15 PPRGR 19

## RESULT 15

S41316  
 coat protein - cucurbit aphid-borne yellows virus  
 C/Species: cucurbit aphid-borne yellows virus  
 C/Date: 19-Mar-1997 #sequence\_revision 06-Jun-1997 #text\_change 20-Sep-1999  
 C/Accession: S41316  
 R:Guillev, H.; Wipf-Scheibel, C.; Richards, K.; Lecocq, H.; Jonard, G.  
 submitted to the EMBL Data Library, December 1993  
 A/Description: Nucleotide sequence of cucurbit aphid-borne yellows virus.  
 A/Reference number: S41313  
 A/Accession: S41316  
 A/Molecule type: genomic RNA  
 A/Residues: 1-199 <GUI>  
 A/Cross-references: EMBL:X76931; NID:G441268; PIDN:CAA54252.1; PID:G441272  
 C/Superfamily: potato leaf roll virus coat protein; potato leaf roll virus coat protein  
 C/Keywords: coat protein  
 F:1-199/Domain: potato leaf roll virus coat protein homology <COP>

Query Match 100.0%; Score 30; DB 2; Length 199;  
 Best Local Similarity 100.0%; Pred. No. 84;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 PPRGR 5  
 |||||  
 Db 36 PPRGR 40  
 Search completed: April 6, 2004, 16:16:55  
 Job time : 5.60748 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:31:19 ; Search time 3.2243 Seconds  
(without alignments)  
80.746 Million cell updates/sec

Title: US-10-009-709-12  
Perfect score: 30  
Sequence: 1 PPRGR 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues  
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	101	VE4 HPV41	P27553 human papillomavirus type 41
2	30	100.0	108	VIB3_AGRU	P05352 agrobacterium tumefaciens (strain 15955)
3	30	100.0	166	PRPC_HUMAN	P02810 homo sapien
4	30	100.0	198	YF91_MYCLE	Q49626 mycobacteri
5	30	100.0	202	CTD6_HUMAN	Q02277 mycobacteri
6	30	100.0	272	YMA1_MYCBO	Q02277 mycobacteri
7	30	100.0	292	RM19_MOUSE	Q9d338 mus musculu
8	30	100.0	335	WMAT_RINDR	P41358 rinderpest
9	30	100.0	337	TRPD_HALVO	P52562 halobacteri
10	30	100.0	350	VGLI_PVRRI	P07846 pseudorabie
11	30	100.0	365	WN9A_HUMAN	O14904 homo sapien
12	30	100.0	365	WN9A_MOUSE	Q8r5m2 mus musculu
13	30	100.0	377	TEX9_CABEL	Q22289 caenorhabdi
14	30	100.0	474	CBLC_HUMAN	Q9ulv8 homo sapien
15	30	100.0	504	V4JA_RHLN	P55501 rhizobium s
16	30	100.0	576	Z384_HUMAN	Q8t568 homo sapien
17	30	100.0	579	Z384_RAT	Q9edj4 rattus norv
18	30	100.0	633	ROR_HUMAN	Q43390 homo sapien
19	30	100.0	641	PK1_STRTO	Q9kig4 streptomyce
20	30	100.0	662	DNK1_SYNPX	Q7u6r7 synchococc
21	30	100.0	687	WRK2_ARATH	Q9fg77 arabidopsis
22	30	100.0	861	GLND_FALSO	Q8x2j3 ralstonia s
23	30	100.0	888	TWC2_MOUSE	Q8r4p4 mus musculu
24	30	100.0	1056	Z451_MOUSE	Q8c0p7 mus musculu
25	30	100.0	1168	MYSC_ACACA	P10569 acanthamoeb
26	30	100.0	1311	FNMR2_HUMAN	P51816 homo sapien
27	30	100.0	1336	W146_HUMAN	Q9c0j8 homo sapien
28	30	100.0	1398	MHP1_YEAST	P43638 saccharomyc
29	30	100.0	1429	NOS1_MOUSE	Q9z0j4 mus musculu
30	30	100.0	1434	NOS1_HUMAN	P29475 homo sapien
31	30	100.0	1553	TY2A_CHICK	Q42130 gallus gall
32	30	100.0	2114	MY9B_MOUSE	Q9qy06 mus musculu
33	27	90.0	99	MDCC_PSSSM	Q87V56 pseudomonas

RESULT 1  
VE4 HPV41  
ID VE4 HPV41 STANDARD; PRT; 101 AA.  
AC P27553;  
DT 01-AUG-1992 (Rel. 23, Created)  
DT 01-OCT-1993 (Rel. 27, Last sequence update)  
DT 01-OCT-1993 (Rel. 27, Last annotation update)  
DE Probable E4 protein.  
OS Human papillomavirus type 41.  
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;  
OC Papillomavirus.  
OX NCBI\_TaxID=10589;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91253264; PubMed=1645904;  
RA Hirt L., Hirsch-Behnman A., de Villiers E.M.;  
RT "Nucleotide sequence of human papillomavirus (HPV) type 41: an  
RT unusual HPV type without a typical E2 binding site consensus  
RT sequence.";  
RL Virus Res. 18:179-190(1991).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL; X56147; CAA39616.1; -  
DR PIR; E43550; E43550.  
KW Early protein.  
SQ SEQUENCE 101 AA; 11912 MW; D9P7734FF9268E3E CRC64;

Query Match 100.0%; Score 30; DB 1; Length 101;  
Best Local Similarity 100.0%; Pred. No. 19;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
Db 6 PPRGR 10

RESULT 2  
VIB3\_AGRU  
ID VIB3\_AGRU STANDARD; PRT; 108 AA.  
AC P05352;  
DT 01-NOV-1988 (Rel. 09, Created)  
DT 01-NOV-1988 (Rel. 09, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE VIB3 protein.

OS Agrobacterium tumefaciens, and  
OS Agrobacterium tumefaciens (strain 15955).  
OG Plasmid pTiA6, and Plasmid pTi15955.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;

Q82dm3 streptomyce  
P72403 streptomyce  
Q9y9d8 aerobium p  
P23684 nevea brasl  
Q9a724 caulobacter  
P23290 xenopus lae  
P13319 bacterioph  
P23163 african swi  
Q13454 homo sapien  
Q9nyto homo sapien  
Q9wv52 mus musculu  
P70105 cavia porce

ALIGNMENTS

```

OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=358, 190386;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=15955; PLASMID=PT115955;
RX MEDLINE=88247765; PubMed=2837739;
RA Thompson D.V., Melchers L.S., Idler K.B., Shilperoort R.A.,
RA Hooykaas P.J.J.;
RT "Analysis of the complete nucleotide sequence of the Agrobacterium
RT tumefaciens virB operon.";
RL Nucleic Acids Res. 16:4621-4636(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC PLASMID=PT1A6;
RX MEDLINE=88186901; PubMed=3281947;
RA Ward J.E., Akiyoshi D.E., Regier D., Datta A., Gordon M.P.,
RA Nester E.W.;
RT "Characterization of the virB operon from an Agrobacterium
RT tumefaciens Ti plasmid.";
RL J. Biol. Chem. 263:5804-5814(1988).
CC -!- FUNCTION: VirB proteins are suggested to act at the bacterial
CC surface and there play an important role in directing T-DNA
CC transfer to plant cells.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X06826; CRA29974.1; -.
DR EMBL; J03216; AAA88647.1; -.
DR PIR; S00779; B3AG55.
DR InterPro; IP3007792; VirB3.
DR Pfam; PF05103; VirB3; 1.
KW Crown gall tumor; Plasmid.
SQ SEQUENCE 108 AA; 11760 MW; 3D0CCB19ECA105D CRC64;

Query Match 100.0%; Score 30; DB 1; Length 108;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5
DB 101 PPRGR 105

RESULT 3
PRPC_HUMAN
ID PRPC_HUMAN STANDARD; PRT; 166 AA.
AC P02810;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-
DE 3) (PRP-2/PRP-4) (PIF-P/PIF-S) (Protein A/protein C) [Contains:
DE Peptide P-C].
GN PRH1 AND PRH2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ALLELES PRH1-4 AND PRH2-1).
RX MEDLINE=86196106; PubMed=3009472;
RA Kim H.-S., Maeda N.;
RT "Structures of two HaeIII-type genes in the human salivary
RT proline-rich protein multigene family.";
RL J. Biol. Chem. 261:6712-6718(1986).
RN [2]
RP SEQUENCE FROM N.A. (ALLELES PRH2-1 AND PRH1-PIF).

```

```

RX MEDLINE=85289325; PubMed=2993301;
RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;
RT "Differential RNA splicing and post-translational cleavages in the
RT human salivary proline-rich protein gene system.";
RL J. Biol. Chem. 260:11123-11130(1985).
RN [3]
RP SEQUENCE OF 17-166 (PRP-1 TO PRP-4; PIF-F AND PIF-S).
RX MEDLINE=89061650; PubMed=3196309;
RA Hay D.I., Bennick A., Schlesinger D.H., Minaguchi K.,
RA Madapallimattam G., Schluckebier S.K.;
RT "The primary structures of six human salivary acidic proline-rich
RT proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-s and PIF-f).";
RL Biochem. J. 255:15-21(1988).
RN [4]
RP SEQUENCE OF 17-166 FROM N.A. (PA ALLELE).
RX MEDLINE=88074309; PubMed=3687941;
RA Azen E.A., Kim H.-S., Goodman P., Flynn S., Maeda N.;
RT "Alleles at the PRH1 locus coding for the human salivary-acidic
RT proline-rich proteins Pa, Db, and PIF.";
RL Am. J. Hum. Genet. 41:1035-1047(1987).
RN [5]
RP SEQUENCE OF 17-166 (PRP-2).
RX MEDLINE=86222916; PubMed=3710693;
RA Schlesinger D.H., Hay D.I.;
RT "Complete covalent structure of a proline-rich phosphoprotein, PRP-2,
RT an inhibitor of calcium phosphate crystal growth from human parotid
RT saliva.";
RL Int. J. Pept. Protein Res. 27:373-379(1986).
RN [6]
RP SEQUENCE OF 17-166 (PROTEIN C).
RX MEDLINE=80204368; PubMed=7380845;
RA Wong R.S.C., Bennick A.;
RT "The primary structure of a salivary calcium-binding proline-rich
RT phosphoprotein (protein C), a possible precursor of a related
RT salivary protein A.";
RL J. Biol. Chem. 255:5943-5948(1980).
RN [7]
RP SEQUENCE OF 17-46 (PROTEIN C).
RX MEDLINE=81191179; PubMed=7228490;
RA Schlesinger D.H., Hay D.I.;
RT "Primary structure of the active tryptic fragments of human and
RT monkey salivary anionic proline-rich proteins.";
RL Int. J. Pept. Protein Res. 17:34-41(1981).
RN [8]
RP SEQUENCE OF 17-122 (PROTEIN A).
RX MEDLINE=79173237; PubMed=438215;
RA Wong R.S.C., Hofmann T., Bennick A.;
RT "The complete primary structure of a proline-rich phosphoprotein from
RT human saliva.";
RL J. Biol. Chem. 254:4800-4808(1979).
RN [9]
RP SEQUENCE OF 17-122 (PROTEIN A).
RA Schlesinger D.H., Hay D.I.;
RT "Complete primary structure of a proline-rich phosphoprotein (PRP-4),
RT a potent inhibitor of calcium phosphate precipitation in human parotid
RT saliva.";
RL (in) Gross E., Meienhofer J. (eds.);
RL Peptides: Structure and Biological Function (Proceedings of the 6th
RL American peptide symposium), pp.133-136, Pierce Chemical Co.,
RL Rockford IL. (1979).
RN [10]
RP SEQUENCE OF 123-166 (PEPTIDE P-C).
RX MEDLINE=80227634; PubMed=7390979;
RA Isemura S., Saitoh E., Sanada K.;
RT "The amino acid sequence of a salivary proline-rich peptide, P-C, and
RT its relation to a salivary proline-rich phosphoprotein, protein C.";
RL J. Biochem. 87:1071-1077(1980).
RN [11]
RP VARIANT PRH2-3 LYS-163.
RA Azen E.A.;
RT "A frequent mutation in the acidic proline-rich protein gene, PRH2,
RT causing a Q147K change closely adjacent to the bacterial binding
RT domain of the cognate salivary PRP (Pr1') in Afro-Americans.";

```

```

RL Hum. Mutat. 12:72-72(1998).
CC -!- FUNCTION: PRP's act as highly potent inhibitors of crystal growth
CC of calcium phosphates. They provide a protective and reparative
CC environment for dental enamel which is important for the integrity
CC of the teeth.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- PTM: Proteolytically cleaved; PRP-2, PRP-1, and PIF-S yield PRP-4,
CC PRP-3 (protein A), and PIF-F, respectively.
CC -!- POLYMORPHISM: Allele PRH1-4 is also known as PRH1 DA allele;
CC allele PRH2-1 is also known as PR1 or protein C; allele PRH2-3 is
CC also known as PR1.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K03202; AAA60183.1; -
DR EMBL; K03203; AAA60184.1; -
DR EMBL; M13057; AAA98807.1; -
DR EMBL; M13058; AAA98808.1; -
DR Genew; HGNC:9366; PRH1.
DR Genew; HGNC:9367; PRH2.
DR MIM; 168730; -
DR MIM; 168790; -
DR MIM; 168710; -
DR GO; GO:0005615; C:extracellular space; TAS.
DR Repeat; Parotid gland; Phosphorylation; Signal; Polymorphism;
KW Pyroglutamate carboxylic acid.
KW SIGNAL
FT CHAIN 1 16
FT CHAIN 17 166
FT CHAIN 17 122
FT CHAIN 17 122
FT CHAIN 123 166
FT DOMAIN 17 46
FT MOD RES 17 17
FT MOD RES 24 24
FT MOD RES 38 38
FT VARIANT 20 20
FT VARIANT 66 66
FT VARIANT 163 163
FT CONFLICT 41 41
FT CONFLICT 166 AA; 17017 MW; A7DF62BF94E3C3EF CRC64;
SQ SEQUENCE 166 AA; 17017 MW; 17017 MW; A7DF62BF94E3C3EF CRC64;

Query Match 100.0%; Score 30; DB 1; Length 166;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PPRGR 5
Db 120 PPRGR 124

RESULT 4
ID YF91 MYCLE STANDARD; PRT; 198 AA.
AC Q49626;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein ML1222.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.

```

```

OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RN STRAIN=TN;
RC MEDLINE=21129732; PubMed=11234002;
RA Cole S.T., Eigelmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Macleod C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Ruster S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: STRONG, TO M.TUBERCULOSIS RV1591.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00010; AAA17073.1; -
DR EMBL; AL583921; CAC31603.1; -
DR PIR; S72709; S72709.
DR Leproma; ML1222; -
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSEM 20 40 POTENTIAL.
FT TRANSEM 70 90 POTENTIAL.
FT TRANSEM 107 127 POTENTIAL.
FT TRANSEM 164 184 POTENTIAL.
SQ SEQUENCE 198 AA; 20633 MW; 65B0422C4B584EBE CRC64;

Query Match 100.0%; Score 30; DB 1; Length 198;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PPRGR 5
Db 15 PPRGR 19

RESULT 5
CTD6 HUMAN STANDARD; PRT; 202 AA.
ID CTD6 HUMAN STANDARD; PRT; 202 AA.
AC Q9BYL1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein C20orf136.
GN C20ORF136.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
RA Beasley C.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,

```

```

RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Cuisinon A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graffam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levasalho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
RA Suze C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.,
RT "The DNA sequence and comparative analysis of human chromosome 20."
RL Nature 414:865-871(2001).
CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL118506; CAC28315.1; -
DR Genew; HGNC:16129; C2orf136.
DR InterPro; IPR001660; SAM.
DR Pfam; PF00536; SAM; 1.
DR SMART; SM00454; SAM; 1.
DR PROSITE; PS50105; SAM_DOMAIN; 1.
KW Hypothetical protein.
FT DOMAIN 118 184
FT SEQUENCE 202 AA; 22770 MW; DF2678F090A3E946 CRC64;
SQ
Query Match 100.0%; Score 30; DB 1; Length 202;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PPRGR 5
DB 11 PPRGR 15
RESULT 6
ID YMA1 MYCBO STANDARD; PRT; 272 AA.
AC Q02277.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative mycocerosyl transferase in mas 5' region.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BCG;
RX MEDLINE=92406887; PubMed=1527058;
RA Mathur M., Kolatukudy P.E.;
RT "Molecular cloning and sequencing of the gene for mycocerosic acid
RT synthase, a novel fatty acid elongating multifunctional enzyme, from
RT Mycobacterium tuberculosis var. bovis Bacillus Calmette-Guerin."
RL J. Biol. Chem. 267:19388-19395(1992).
CC -!- FUNCTION: PROBABLY TRANSFERS THE MYCOCEROSYL GROUP FROM
CC MYCOCEROSIC ACID SYNTHASE TO THE HYDROXYL GROUP OF

```

```

CC PHENOLPHTHIOEROL.
CC -!- SIMILARITY: TO M.LEPRAE MAB.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M95808; AAA25368.1; -
DR PIR; A44110; A44110.
KW Hypothetical protein; Transferase; Transmembrane.
SQ SEQUENCE 272 AA; 29827 MW; C31A93100E769229 CRC64;
Query Match 100.0%; Score 30; DB 1; Length 272;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 PPRGR 5
DB 183 PPRGR 187
RESULT 7
ID RM19 MOUSE STANDARD; PRT; 292 AA.
AC Q9D338; O8R1R0.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 60S ribosomal protein L19, mitochondrial precursor (L19mt).
GN MRPL19.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Colon;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.W., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Kamiya M., Lee N.H.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Sakamoto N.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection."
RL Nature 409:685-690(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore S.I., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,

```



```

RA Raba S.S., Loquellano N.A., Peters G.J., Abranson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gnaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.O., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA Schnerch A., Schein J.B., Jones S.J.M., Marra M.A.,
RT human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -!- SUBCELLULAR LOCATION: Mitochondrial.
CC -!- SIMILARITY: Belongs to the L19p family of ribosomal proteins
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AK018508; BAB31245.1; -
DR EMBL; BC020315; AAH20315.1; -
DR EMBL; BC043921; AAH43921.1; -
DR MGD; MGI:1926274; Mrlp119.
DR InterPro; IPR001857; Ribosomal L19.
DR Pfam; PF01245; Ribosomal L19; I.
DR PRINTS; PR00861; Ribosomal L19.
DR PRODOM; PD002979; Ribosomal L19; 1.
DR PROSITE; PS01015; RIBOSOMAL L19; FALSE NEG.
KW Ribosomal protein; Mitochondrion; Transit peptide.
FT CHAIN ? 292 60S RIBOSOMAL PROTEIN L19.
FT TRANSIT ? 292 60S RIBOSOMAL PROTEIN L19.
SQ SEQUENCE 292 AA; 33578 MW; 2D85C99D97F5A59BF5 CRC64;

Query Match 100.0%; Score 30; DB 1; Length 292;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PPRGR 5
Db 82 PPRGR 86

RESULT 8
VMA1_RINDR
ID VMA1_RINDR STANDARD; PRT; 335 AA.
AC P41358;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Matrix protein.
GN M.
OS Rinderpest virus (strain RBOK) (RDV).
OC Viruses; ssRNA negative-strand viruses; Mononegavirales;
OC Paramyxoviridae; Paramyxovirinae; Morbillivirus.
OX NCBI_TaxID=36409;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95205080; PubMed=7897350;
RA Baron M.D., Barrett T.;
RT "Sequencing and analysis of the nucleocapsid (N) and polymerase (L)
RT genes and the terminal extragenic domains of the vaccine strain of
RT rinderpest virus."
RL J. Gen. Virol. 76:593-602(1995).
CC -!- FUNCTION: THE M PROTEIN HAS A CRUCIAL ROLE IN VIRUS ASSEMBLY
CC AND INTERACTS WITH THE RNP COMPLEX AS WELL AS WITH THE VIRAL
CC MEMBRANE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z30697; CAA83180.1; -
DR PIR; S47304; S47304.
DR InterPro; IPR000982; Matrix.
DR Pfam; PF00661; Matrix; 1.
DR PRODOM; PD000741; Matrix; 1.
KW Matrix protein; Envelope protein.
SQ SEQUENCE 335 AA; 37610 MW; B762DF12E2A06ECC CRC64;

Query Match 100.0%; Score 30; DB 1; Length 335;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PPRGR 5
Db 67 PPRGR 71

RESULT 9
TRPD_HALVO
ID TRPD_HALVO STANDARD; PRT; 337 AA.
AC P52562;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Anthranilate phosphoribosyltransferase (EC 2.4.2.18).
GN TRPD.
OS Halobacterium volcanii (Haloflex volcanii).
OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
OC Halobacteriaceae; Haloflex.
OX NCBI_TaxID=2246;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=WFD11;
RX MEDLINE=92165748; PubMed=1537810;
RA Lam W.L., Logan S.M., Doolittle W.F.;
RT "Genes for tryptophan biosynthesis in the halophilic archaeobacterium
RT Haloflex volcanii: the trpDREG cluster."
RL J. Bacteriol. 174:1694-1697(1992).
CC -!- CATALYTIC ACTIVITY: Anthranilate + phosphoribosyl-diphosphate =
CC N-5'-phosphoribosyl-anthranilate + diphosphate.
CC -!- PATHWAY: Tryptophan biosynthesis; second step.
CC -!- SIMILARITY: Belongs to the anthranilate phosphoribosyltransferase
CC family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M83788; AAA73175.1; -
DR HAMAP; MF 00211; -; 1.
DR InterPro; IPR005940; Ant. phosho trans.
DR InterPro; IPR000312; Glyco_transf_3.
DR Pfam; PF02885; Glycos_transf_3N; 1.
DR Pfam; PF00591; Glycos_transf_3; 1.
DR PRODOM; PD001864; Glyco_transf_3; 1.
DR TIGRFAMs; TIGR01245; trpD; 1.
KW Tryptophan biosynthesis; Transferase; Glycosyltransferase.
SQ SEQUENCE 337 AA; 34551 MW; 13ECFAF2A0AE941E CRC64;

Query Match 100.0%; Score 30; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY      1 PPRGR 5
DB      22 PPRGR 26

RESULT 10
VGLI_PVR1
ID_VGLI_PVR1 STANDARD; PRT; 350 AA.
AC P07646;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-APR-1988 (Rel. 07, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Glycoprotein GP63 precursor.
OS Pseudorabies virus (strain Rice) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=10350;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=86308235; PubMed=3018284;
RA Petrovskis E.A., Timmins J.G., Post L.E.;
RT "Use of lambda g11 to isolate genes for two pseudorabies virus
RT glycoproteins with homology to herpes simplex virus and varicella-
RT zoster virus glycoproteins."
RL J. Virol. 60:185-193(1986).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: TO OTHER HERPESVIRUSES GLYCOPROTEINS I, AND TO VZV
CC GP1V.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; M14336; AAC35204.1; -
DR PIR; A29012; VGBE63.
DR InterPro; IPR002874; Herpes_G1.
DR Pfam; PF01688; Herpes_G1; 1.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 23 OR 27 (POTENTIAL).
FT CHAIN 24 350 GLYCOPROTEIN GP63.
FT DOMAIN 24 285 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 286 308 POTENTIAL.
FT DOMAIN 309 350 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 56 56 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 73 73 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 153 153 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 262 262 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 350 AA; 36773 MW; 66AF2229EC31BEDA CRC64;

Query Match 100.0%; Score 30; DB 1; Length 350;
Best Local Similarity 100.0%; Pred. No. 67;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PPRGR 5
DB      166 PPRGR 170

RESULT 11
WN9A_HUMAN
ID WN9A_HUMAN STANDARD; PRT; 365 AA.
AC O14904; Q96S50;
DT 15-JUL-1998 (Rel. 36, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Wnt-9a protein precursor (Wnt-14).

```

```

GN WNT9A OR WNT14.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
MEDLINE=21308441; PubMed=11414706;
RA Saitoh T., Hirai M., Kato M.;
RT "Molecular cloning and characterization of WNT3a and WNT14 clustered
RT in human chromosome 1q42 region."
RL Biochem. Biophys. Res. Commun. 284:1168-1175(2001).
RN [2]
SEQUENCE OF 221-343 FROM N.A.
MEDLINE=98110581; PubMed=9441749;
RA Bergstein I., Eisenberg L.M., Bhallerao J., Jenkins N.A.;
RA Copeland N.G., Osborne M.P., Bowcock A.M., Brown A.M.C.;
RT "Isolation of two novel WNT genes, WNT14 and WNT15, one of which
RT (WNT15) is closely linked to WNT3 on human chromosome 17q21."
RL Genomics 46:450-458(1997).
CC -!- FUNCTION: Ligand for members of the frizzled family of seven
CC transmembrane receptors. Probable developmental protein. May be a
CC signaling molecule which affects the development of discrete
CC regions of tissues. Is likely to signal over only few cell
CC diameters (by similarity).
CC -!- SUBCELLULAR LOCATION: Possibly secreted and associates with the
CC extracellular matrix.
CC -!- SIMILARITY: Belongs to the Wnt family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; AB060283; BAB61051.1; -
DR EMBL; AF028702; AAC39550.1; -
DR Genew; HGNC:12778; WNT9A.
DR MIM; 602863; -
DR GO; GO:0005576; C:extracellular; NAS.
DR GO; GO:0005201; F:extracellular matrix structural constituent; NAS.
DR GO; GO:0007267; P:cell-cell signaling; NAS.
DR GO; GO:0007275; P:development; NAS.
DR InterPro; IPR005816; Wnt_grthfactor.
DR Pfam; PF00110; wnt; 1.
DR PRINTS; PR01349; WNTPROTEIN.
DR SMART; SM00097; WNT1; 1.
DR PROSITE; PS00246; WNT1; 1.
KW Wnt signalling pathway; Developmental protein; Glycoprotein; Signal.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 365 WNT-9A PROTEIN.
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 365 AA; 40320 MW; 1E1284D744C6A9B2 CRC64;

Query Match 100.0%; Score 30; DB 1; Length 365;
Best Local Similarity 100.0%; Pred. No. 70;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 PPRGR 5
DB      269 PPRGR 273

RESULT 12
WN9A_MOUSE
ID WN9A_MOUSE STANDARD; PRT; 365 AA.
AC Q8R5M2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)

```



CC Name-Short;  
CC Isoid=Q9ULV8-2; Sequence=VSP\_005732;  
CC TISSUE SPECIFICITY: Ubiquitous.  
CC !- DOMAIN: The N-terminus is composed of the phosphotyrosine binding  
CC (PTB) domain, a short linker region and the RING-type zinc finger.  
CC The PTB domain, which is also called TKB (tyrosine kinase binding)  
CC domain, is composed of three different subdomains: a four-helix  
CC bundle (4H), a calcium-binding EF hand and a divergent SH2 domain.  
CC !- DOMAIN: The RING-type zinc finger domain mediates binding to an E2  
CC ubiquitin-conjugating enzyme (By similarity).  
CC !- PTM: Phosphorylated on tyrosines by EGFR.  
CC !- MISCELLANEOUS: This protein has one functional calcium-binding  
CC site (By similarity).  
CC !- SIMILARITY: Contains 2 EF-hand-like calcium-binding domains.  
CC !- SIMILARITY: Contains 1 RING-type zinc finger.  
CC !- SIMILARITY: Contains 1 SH2 domain.  
CC !- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.; CBLCID194.html".  
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/CBLCID194.html".  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AB028645; BAA86298.1; -  
CC EMBL; AF117646; AAD34341.1; -  
CC EMBL; AF117647; AAD34342.1; -  
CC HSSP; P22661; 1B47.  
CC Genew; HGNC:15961; CBLC.  
CC GO; GO:0005515; F:protein binding; TAS.  
CC GO; GO:0008270; F:zinc ion binding; TAS.  
CC GO; GO:0007175; P:negative regulation of EGF receptor activity; TAS.  
CC GO; GO:0006468; P:protein amino acid phosphorylation; TAS.  
CC InterPro; IPR003153; Cbl\_N.  
CC InterPro; IPR000380; SH2.  
CC InterPro; IPR001841; Znf\_ring.  
CC Pfam; PF02761; Cbl\_N2; 1.  
CC Pfam; PF02762; Cbl\_N3; 1.  
CC Pfam; PF02262; Cbl\_N; 1.  
CC Pfam; PF00097; Zf-G3HC4; 1.  
CC SMART; SM00184; RING; 1.  
CC SMART; SM00252; SH2; 1.  
CC PROSITE; PS00518; ZF\_RING\_1; 1.  
CC PROSITE; PS00089; ZF\_RING\_2; 1.  
CC Nuclear protein; Zinc-finger; Phosphorylation; Alternative splicing;  
CC SH2 domain; Calcium-binding.  
CC PTB.  
CC DOMAIN 8 327  
CC DOMAIN 8 147  
CC DOMAIN 182 190  
CC EF-HAND LIKE 1.  
CC DOMAIN 199 210  
CC EF-HAND LIKE 2.  
CC SH2.  
CC DOMAIN 237 311  
CC LINKER.  
CC DOMAIN 312 350  
CC RING-TYPE.  
CC ZN\_FING 351 390  
CC NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).  
CC DOMAIN 89 92  
CC PHOSPHOTYROSINE (BY SIMILARITY).  
CC BINDING 264 284  
CC Missing (in isoform Short).  
CC VARSPLIC 261 306  
CC /FTID=VSP\_005732.  
CC T -> N (IN REF. 2).  
CC CONFLICT 234 234  
CC SEQUENCE 474 AA; 52468 MW; 91013DDF12828242 CRC64;

Query Match 100.0%; Score 30; DB 1; Length 474;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PPRGR 5  
DB 99 PPRGR 103

Y4JA\_RHISN STANDARD; PRT; 504 AA.  
ID Y4JA\_RHISN  
AC P55501; 1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE Hypothetical 57.2 kDa protein Y4JA/Y4NE/Y4SE.  
GN Y4JA AND Y4NE AND Y4SE.  
OS Rhizobium sp. (strain NGR234).  
OG Plasmid sym pNGR234a.  
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.  
OX NCBI\_TaxID=394;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=97305956; PubMed=9163424;  
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,  
RA Perret X.; basis of symbiosis between Rhizobium and legumes.";  
RT "Molecular basis of symbiosis between Rhizobium and legumes.";  
RL Nature 387:394-401(1997).  
CC !- SIMILARITY: VERY LOW SIMILARITY TO THE IS21/IS408/IS1162 FAMILY OF  
CC TRANSPOSASES.  
CC !- SIMILARITY: STRONG, TO FZ4.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AE000079; AAB91713.1; -  
CC EMBL; AE000086; AAB91785.1; -  
CC EMBL; AE000095; AAB91845.1; -  
CC InterPro; IPR001584; Rve.  
CC Pfam; PF00665; rve; 1.  
CC Hypothetical protein; Plasmid.  
CC SEQUENCE 504 AA; 57228 MW; 63E31E487DD4CB87 CRC64;  
Query Match 100.0%; Score 30; DB 1; Length 504;  
Best Local Similarity 100.0%; Pred. No. 98;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PPRGR 5  
DB 52 PPRGR 56

Search completed: April 6, 2004, 16:08:01  
Job time : 3.243 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:51:34 ; Search time 18.0941 Seconds  
(without alignments)  
87.236 Million cell updates/sec

Title: US-10-009-709-12

Perfect score: 30

Sequence: 1 PPRGR 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp archaea.\*

2: sp bacteria.\*

3: sp fungi.\*

4: sp human.\*

5: sp invertebrate.\*

6: sp mammal.\*

7: sp mhc.\*

8: sp organelle.\*

9: sp phage.\*

10: sp plant.\*

11: sp rodent.\*

12: sp virus.\*

13: sp vertebrate.\*

14: sp unclassified.\*

15: sp rvirus.\*

16: sp bacteriaph.\*

17: sp archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	36	11	Q61356 mus musculus
2	30	100.0	40	4	Q14069 Q14069 homo sapien
3	30	100.0	62	11	Q63326 Q63326 rattus norv
4	30	100.0	63	16	Q7UXT5 Q7UXT5 rhodopirell
5	30	100.0	79	17	Q8ZTW7 Q8ZTW7 pyrobaculum
6	30	100.0	82	16	Q7UPE8 Q7UPE8 rhodopirell
7	30	100.0	105	16	Q82FW1 Q82FW1 streptomyce
8	30	100.0	108	2	Q9R2G3 Q9R2G3 agrobacteri
9	30	100.0	118	16	Q882E4 Q882E4 pseudomonas
10	30	100.0	126	12	Q91TR0 Q91TR0 tupaia herp
11	30	100.0	127	10	Q8W390 Q8W390 oryza sativ
12	30	100.0	143	11	Q91Y15 Q91Y15 mus musculu
13	30	100.0	143	17	Q9YDA0 Q9YDA0 aeropyrum p
14	30	100.0	150	17	Q9YDY3 Q9YDY3 aeropyrum p
15	30	100.0	154	16	Q7WJM3 Q7WJM3 bordetella
16	30	100.0	154	16	Q7WAH6 Q7WAH6 bordetella

17	30	100.0	155	4	Q8WVB8 Q8WVB8 homo sapien
18	30	100.0	155	4	Q9HA73 Q9HA73 homo sapien
19	30	100.0	155	5	Q9W199 Q9W199 drosophila
20	30	100.0	155	6	Q77618 Q77618 monodelphis
21	30	100.0	156	12	Q99AR6 Q99AR6 tt virus. o
22	30	100.0	161	4	Q8NH35 Q8NH35 homo sapien
23	30	100.0	163	2	Q9LB07 Q9LB07 vibrio para
24	30	100.0	163	16	Q87MK7 Q87MK7 vibrio para
25	30	100.0	164	10	Q9FUL3 Q9FUL3 oryza sativ
26	30	100.0	165	10	Q949E3 Q949E3 oryza sativ
27	30	100.0	166	10	Q8RZK8 Q8RZK8 oryza sativ
28	30	100.0	166	16	Q89D41 Q89D41 bradyrhizob
29	30	100.0	167	16	Q8UI60 Q8UI60 agrobacteri
30	30	100.0	168	4	Q9NSY3 Q9NSY3 homo sapien
31	30	100.0	174	16	Q82312 Q82312 streptomyce
32	30	100.0	176	16	Q82P20 Q82P20 streptomyce
33	30	100.0	177	16	Q8KI47 Q8KI47 streptomyce
34	30	100.0	178	17	Q9UZ77 Q9UZ77 pyrococcus
35	30	100.0	186	16	Q82GY4 Q82GY4 streptomyce
36	30	100.0	188	10	Q84RX3 Q84RX3 oryza sativ
37	30	100.0	189	10	Q9FK53 Q9FK53 arabidopsis
38	30	100.0	190	11	Q811B5 Q811B5 mus musculu
39	30	100.0	197	16	Q7U0S9 Q7U0S9 mycobacteri
40	30	100.0	199	12	Q85970 Q85970 cucumber ye
41	30	100.0	200	12	Q66619 Q66619 equine herp
42	30	100.0	204	16	Q9KR85 Q9KR85 streptomyce
43	30	100.0	206	10	Q7XQ10 Q7XQ10 oryza sativ
44	30	100.0	211	13	Q9PWP2 Q9PWP2 gallus gall
45	30	100.0	213	10	Q7XI49 Q7XI49 oryza sativ

## ALIGNMENTS

### RESULT 1

Q61356	PRELIMINARY;	PRT;	36 AA.
ID Q61356			
AC Q61356			
DT 01-NOV-1996	(TREMBLrel. 01, Created)		
DT 01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT 01-JUN-2003	(TREMBLrel. 24, Last annotation update)		
DE BMP-1	protein (Fragment).		
GN BMP1 OR BMP-1.			
OS Mus musculus (Mouse).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX NCBI_TaxID=10090;			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE=94229342; PubMed=8174772;			
RA Fukagawa M., Suzuki N., Hogan B.L., Jones C.M.;			
RT "Embryonic expression of mouse bone morphogenetic protein-1 (BMP-1),			
RT which is related to the Drosophila dorsoventral gene tolloid and			
RT encodes a putative atactin metalloendopeptidase.";			
RL Dev. Biol. 163:175-183(1994).			
RN [2]			
RP SEQUENCE FROM N.A.			
RX MEDLINE=95096114; PubMed=7798260;			
RA Takahara K., Lyons G.E., Greenspan D.S.;			
RT "Bone morphogenetic protein-1 and a mammalian tolloid homologue (mtld)			
RT are encoded by alternatively spliced transcripts which are			
RT differentially expressed in some tissues.";			
RL J. Biol. Chem. 269:32572-32578(1994).			
DR EMBL; L35280; AAC42043.1; -			
DR PIR; C58788; C58788.			
DR MGD; MGI:88176; Bmp1.			
FT NON TER			
SQ SEQUENCE	36 AA; 4342 MW; 0972E9E7D0469768 CRC64;		

Query Match 100.0%; Score 30; DB 11; Length 36;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
 |||||  
 Db 16 PPRGR 20

## RESULT 2

Q14069 Q14069 PRELIMINARY; PRT; 40 AA.  
 AC Q14069;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
 DE (Clone E18) Gene from CpG-enriched DNA (Fragment).  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95135425; PubMed=7833926;  
 RA Bonaldo M.F., Yu M.T., Jelenc P., Brown S., Su L., Lawton L.,  
 RA Deaven L., Efstratiadis A., Warburton D., Soares M.B.,  
 RT "Selection of cDNAs using chromosome-specific genomic clones:  
 RT Application to human chromosome 13."  
 RL Hum. Mol. Genet. 3:1663-1673(1994).  
 DR EMBL; L33990; AAA74367.1;  
 FT NON TER 1  
 FT NON TER 40 40  
 SQ SEQUENCE 40 AA; 4466 MW; 287514474C472C38 CRC64;

Query Match 100.0%; Score 30; DB 4; Length 40;  
 Best Local Similarity 100.0%; Pred. No. 46;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
 |||||  
 Db 20 PPRGR 24

## RESULT 3

Q63326 Q63326 PRELIMINARY; PRT; 62 AA.  
 AC Q63326;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Rat malic enzyme (ME) (Fragment).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89096948; PubMed=3211151;  
 RA Morioka H., Tenyson G.E., Nikodem V.M.,  
 RT "Structural and functional analysis of the rat malic enzyme gene  
 RT promoter."  
 RL Mol. Cell. Biol. 8:3542-3545(1988).  
 RN [2]  
 RP SEQUENCE OF 1-3 FROM N.A.  
 RX MEDLINE=90237036; PubMed=2332433;  
 RA Petty K.J., Desvergne B., Mitsuhashi T., Nikodem V.M.,  
 RT "Identification of a thyroid hormone response element in the malic  
 RT enzyme gene."  
 RL J. Biol. Chem. 265:7395-7400(1990).  
 DR EMBL; M35258; AAA41564.1;  
 FT NON TER 62 62  
 SQ SEQUENCE 62 AA; 6617 MW; AAC6F60B9F96DD47 CRC64;

Query Match 100.0%; Score 30; DB 11; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 70;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
 |||||  
 Db 46 PPRGR 50

## RESULT 4

Q70XT5 Q70XT5 PRELIMINARY; PRT; 63 AA.  
 AC Q70XT5;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN RB1133.  
 OS Rhodospirillum rubrum.  
 OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
 OC Planctomycetaceae; Pirellula.  
 OX NCBI\_TaxID=117;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=1;  
 RX MEDLINE=22735313; PubMed=12835416;  
 RA Gloeckner F.O., Rube M., Bauer M., Teeling H., Lombardot T.,  
 RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,  
 RA Schlesner H., Amann R., Reinhardt R.,  
 RT "Complete genome sequence of the marine planctomycete Pirellula sp.  
 RT strain 1."  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).  
 DR EMBL; BX294134; CAD71918.1;  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 63 AA; 6873 MW; D14BE2CD57131BD3 CRC64;

Query Match 100.0%; Score 30; DB 16; Length 63;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
 |||||  
 Db 57 PPRGR 61

## RESULT 5

Q82TW7 Q82TW7 PRELIMINARY; PRT; 79 AA.  
 AC Q82TW7;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein PAE3058.  
 GN PAE3058.  
 OS Pyrobaculum aerophilum.  
 OC Archaea; Crenarchaeota; Thermoprotei; Thermoproteales;  
 OC Thermoproteaceae; Pyrobaculum.  
 OX NCBI\_TaxID=13773;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX STRAIN=IM2 / ATCC 51768 / DSM 7523;  
 RX MEDLINE=21664397; PubMed=11792869;  
 RA Fitz-Gibbon S.T., Ladner H., Kim U.-J., Stetter K.O., Simon M.I.,  
 RA Miller J.H.,  
 RT "Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum  
 RT aerophilum."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:984-989(2002).  
 DR EMBL; AB009908; AAL64642.1;  
 DR InterPro; IPR002792; TRAM.  
 DR Pfam; PF01938; TRAM; 1.  
 KW Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 79 AA; 8695 MW; B36D5871B6C30143 CRC64;

Query Match 100.0%; Score 30; DB 17; Length 79;  
 Best Local Similarity 100.0%; Pred. No. 88;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 PPRGR 5
DB 9 PPRGR 13

RESULT 6
ID Q7UPB8 PRELIMINARY; PRT; 82 AA.
AC Q7UPB8;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein.
GN RB6989.
OS Rhodopirellula baltica.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,
RA Schlesner H., Anann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1."
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
DR EMBL; BX294145; CAD75114.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 82 AA; 9143 MW; 5004CB56496A2944 CRC64;

Query Match 100.0%; Score 30; DB 16; Length 82;
Best Local Similarity 100.0%; Pred. No. 91;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5
DB 76 PPRGR 80

RESULT 7
ID Q82FW1 PRELIMINARY; PRT; 105 AA.
AC Q82FW1;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein.
GN SAV4141.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osone T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis; deducing the ability of producing secondary
RT metabolites."
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
MEDLINE=22608306; PubMed=12692562;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis."

```

```

RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL; AP005037; BAC71853.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 105 AA; 11436 MW; 14E905DB7C2CA475 CRC64;

Query Match 100.0%; Score 30; DB 16; Length 105;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5
DB 88 PPRGR 92

RESULT 8
ID Q9R2G3 PRELIMINARY; PRT; 108 AA.
AC Q9R2G3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE VIRB3-VIRULENCE protein.
OS Agrobacterium tumefaciens.
OG Plasmid.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=358;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94010218; PubMed=8405938;
RA Shirasu K., Kado C.I.;
RT "Membrane location of the Ti plasmid VirB proteins involved in the
RT biosynthesis of a pilin-like conjugative structure on Agrobacterium
RT tumefaciens."
RL EMS Microbiol. Lett. 111:287-294(1993).
DR InterPro; IPR007732; VirB3.
DR Pfam; PF05101; VirB3; 1.
SQ SEQUENCE 108 AA; 11760 MW; 06616D618D722BD4 CRC64;

Query Match 100.0%; Score 30; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5
DB 101 PPRGR 105

RESULT 9
ID Q882E4 PRELIMINARY; PRT; 118 AA.
AC Q882E4;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein.
GN PSPT02684.
OS Pseudomonas syringae (pv. tomato).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=323;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=BC3000;
RA Buell R., Joardar V., Khouri H., Fedorova N., Tran B., Russell D.,
RA Berry K., Utterback T., Van Aken S., Feldblyum T., Gwinn M.,
RA Dodson R., DeBoy R., Durkin A., Kolonay J., Madupu R., Daugherty S.,
RA Brinkac L., Beanan M., Haft D., Selengut J., Nelson W., Davidsen T.,
RA White O., Fraser C., Collmer A.;
RT "Complete sequence of Pseudomonas syringae."
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016865; AA056185.1; -.
DR TIGR; PSPT02684; -.

```

KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 118 AA; 12797 MW; A6612B0E3FBC341 CRC64;

Query Match 100.0%; Score 30; DB 16; Length 118;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
Db 113 PPRGR 117

## RESULT 10

Q91TRO PRELIMINARY; PRT; 126 AA.  
AC Q91TRO;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE T32.1.  
DE Tupaia herpesvirus.  
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
OC Betaherpesvirinae.  
OX NCBI\_TaxID=10397;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2;  
RX MEDLINE=21211637; PubMed=11312357;  
RA Bahr U.; Darai G.;  
RT "Analysis and Characterization of the Complete Genome of Tupaia (Tree Shrew) Herpesvirus."  
RL J. Virol. 75:4854-4870(2001).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=2;  
RA Darai G.; Bahr U.;  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF281817; AAK57077.1; -;  
SQ SEQUENCE 126 AA; 14152 MW; CC9DF1191FFEC5FB CRC64;

Query Match 100.0%; Score 30; DB 12; Length 126;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
Db 35 PPRGR 39

## RESULT 11

Q8W390 PRELIMINARY; PRT; 127 AA.  
AC Q8W390;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)  
DE Putative chitinase.  
GN OSUNBA0013008.29.  
OS Oryza sativa (Rice).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzeae; Oryza.  
OX NCBI\_TaxID=4530;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cnv. Nipponbare;  
RA Buell C.R.; Yuan Q.; Ouyang S.; Liu J.; Moffat K.S.; Hill J.N.;  
RA Gansberger K.; Brenner M.; Burgess S.; Hance M.; Shwartsbeyn M.;  
RA Tsitrin T.; Riggs F.; Hsiao J.; Zismann V.; Blunt S.; Pai G.;  
RA Varaken S.E.; Uterback T.R.; Feldblyum T.V.; Kalb E.; Quackenbush J.;  
RA Saizberg S.L.; White O.; Fraser C.M.;  
RT "Oryza sativa chromosome 3 BAC OSUNBA0013008 genomic sequence."  
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC084762; AAL58249.1; -;  
DR Gramene; Q8W390; -;  
SQ SEQUENCE 127 AA; 13877 MW; DC58A644FD437DED CRC64;

Query Match 100.0%; Score 30; DB 10; Length 127;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
Db 75 PPRGR 79

## RESULT 12

Q91YI5 PRELIMINARY; PRT; 143 AA.  
ID Q91YI5;  
AC Q91YI5;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN 4930540G07RIK.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Strausberg R.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC016636; AAH6636.1; -;  
DR MGD; MGI:1922460; 4930540G07RIK.  
DR GO; GO:0003676; F:nucleic acid binding; IEA.  
DR InterPro; IPR00571; Znf\_CCH; 1.  
DR Pfam; PF00642; zf-CCH; 1.  
DR SMART; SM00356; Znf\_C3H1; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 143 AA; 15962 MW; B5668CE32FFCAB99 CRC64;

Query Match 100.0%; Score 30; DB 11; Length 143;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
Db 34 PPRGR 38

## RESULT 13

Q9YDA0 PRELIMINARY; PRT; 143 AA.  
ID Q9YDA0;  
AC Q9YDA0;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein APE1012.  
GN APE1012.  
OS Aeropyrum pernix.  
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;  
OC Desulfurococaceae; Aeropyrum.  
OX NCBI\_TaxID=56636;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K1;  
RX MEDLINE=99310339; PubMed=10382966;  
RA Kwarabavasi Y.; Hino Y.; Horikawa H.; Yamazaki S.; Haikawa Y.;  
RA Jin-no K.; Takahashi M.; Sekine M.; Baba S.-I.; Ankaï A.; Kosugi H.;  
RA Hosoyama A.; Fukui S.; Nagai Y.; Nishijima K.; Nakazawa H.;  
RA Takamiya M.; Masuda S.; Funahashi T.; Tanaka T.; Kudoh Y.;  
RA Yamazaki J.; Kishida N.; Oguchi A.; Aoki K.-I.; Kubota K.;  
RA Nakamura Y.; Nomura N.; Sako Y.; Kikuchi H.;  
RT "Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, Aeropyrum pernix K1.";



```

RL DNA Res. 6:83-101(1999).
DR EMBL; AP000060; BAA79997.1; -.
DR PIR; E72699; E72699.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 143 AA; 14821 MW; 7C7FE14DD1EB4CBE CRC64;

Query Match 100.0%; Score 30; DB 17; Length 143;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5
Db 93 PPRGR 97

RESULT 14
Q9VDY3 PRELIMINARY; PRT; 150 AA.
AC Q9VDY3;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein APE0786.
GN APE0786.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Thermoprotei; Desulfurococcales;
OC Desulfurococaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KL;
RX MEDLINE=99310339; PubMed=10382956;
RA Kawatabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Aikai A., Kosugi H.,
RA Hosoyama A., Fukui S., Negai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix KL."
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000060; BAA79764.1; -.
DR PIR; D72670; D72670.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 150 AA; 16568 MW; 03A6699A3B71CD81 CRC64;

Query Match 100.0%; Score 30; DB 17; Length 150;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5
Db 68 PPRGR 72

RESULT 15
Q7WJM3 PRELIMINARY; PRT; 154 AA.
AC Q7WJM3;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN BB2470.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,

```

```

RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdeno-Tarraga A.M., Temple L., James K., Harris B., Quail M.A.,
RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Feltwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagsels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.,
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica."
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640444; CAE32964.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 154 AA; 17002 MW; 0568DF74ED1B25D2 CRC64;

```

```

Query Match 100.0%; Score 30; DB 16; Length 154;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 PPRGR 5
Db 13 PPRGR 17

```

```

Search completed: April 6, 2004, 16:14:44
Job time : 19.161 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:29:39 ; Search time 27.9907 Seconds  
(without alignments)  
50.472 Million cell updates/sec

Title: US-10-009-709-12

Perfect score: 30  
Sequence: 1 PPRGR 5

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : A\_Geneseq\_29Jan04:\*
- 1: Geneseqp1980s:\*
  - 2: Geneseqp1990s:\*
  - 3: Geneseqp2000s:\*
  - 4: Geneseqp2001s:\*
  - 5: Geneseqp2002s:\*
  - 6: Geneseqp2003as:\*
  - 7: Geneseqp2003bs:\*
  - 8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	5	AAB48782	Aab48782 Human sal
2	30	100.0	6	AAY50279	Aay50279 Neutroph
3	30	100.0	6	AAB48781	Aab48781 Human sal
4	30	100.0	7	AAY50278	Aay50278 Neutroph
5	30	100.0	7	AAB48780	Aab48780 Human sal
6	30	100.0	8	AAB48779	Aab48779 Human sal
7	30	100.0	9	AAB48778	Aab48778 Human sal
8	30	100.0	10	AAB48777	Aab48777 Human sal
9	30	100.0	10	AAG33874	Aag33874 Human com
10	30	100.0	17	AAB48783	Aab48783 Human sal
11	30	100.0	25	AAU01685	Aau01685 Gene 28 h
12	30	100.0	28	AAU41484	Aau41484 Fragment
13	30	100.0	34	AAU00630	Aau00630 Human tel
14	30	100.0	34	AAU00658	Aau00658 Telomerax
15	30	100.0	44	ABP62124	Abp62124 Human sec
16	30	100.0	46	ABP42138	Abp42138 Peptide #
17	30	100.0	46	AAM35944	Aam35944 Peptide #
18	30	100.0	46	ABP25701	Abp25701 Protein #
19	30	100.0	46	AAM75833	Aam75833 Human bon
20	30	100.0	46	AAM33022	Aam33022 Human bra
21	30	100.0	46	ABG37569	Abg37569 Human liv
22	30	100.0	50	AAU45927	Aau45927 Propionib
23	30	100.0	50	ABP80084	Abp80084 N. gonorr
24	30	100.0	50	ABM42446	Abm42446 Propionib
25	30	100.0	52	AAU57391	Aau57391 Propionib

26	30	100.0	52	6	ABP77583	Abp77583 N. gonorr
27	30	100.0	52	6	ABM53910	Abm53910 Propionib
28	30	100.0	53	6	ABP80319	Abp80319 N. gonorr
29	30	100.0	56	4	AAU54781	Aau54781 Propionib
30	30	100.0	56	6	ABM51300	Abm51300 Propionib
31	30	100.0	61	4	AAU54510	Aau54510 Propionib
32	30	100.0	61	6	ABM51029	Abm51029 Propionib
33	30	100.0	62	4	AAU49884	Aau49884 Propionib
34	30	100.0	62	4	ABM15033	Abm15033 Human ner
35	30	100.0	62	6	ABM46403	Abm46403 Propionib
36	30	100.0	63	4	AAU53649	Aau53649 Propionib
37	30	100.0	63	6	ABM50168	Abm50168 Propionib
38	30	100.0	64	5	ABP09178	Abp09178 Human ORF
39	30	100.0	65	4	AAU40624	Aau40624 Propionib
40	30	100.0	65	6	ABM37143	Abm37143 Propionib
41	30	100.0	66	4	AAU67516	Aau67516 Propionib
42	30	100.0	66	6	ABM64035	Abm64035 Propionib
43	30	100.0	68	4	AAU45745	Aau45745 Propionib
44	30	100.0	68	6	ABM42264	Abm42264 Propionib
45	30	100.0	69	4	AAU56685	Aau56685 Propionib

ALIGNMENTS

RESULT 1  
AAB48782  
ID AAB48782 standard; peptide; 5 AA.  
XX  
AC AAB48782;  
XX  
DT 09-MAR-2001 (first entry)  
XX  
DE Human saliva PRP-1 fragment (residues 104-108), SEQ ID NO:12.  
XX  
KW Human; PRP-1; proline-rich protein; saliva; dental caries;  
KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
KW oral bacterium; caries prevention.  
XX  
OS Homo sapiens.  
XX  
PN WO200069890-A1.  
XX  
PD 23-NOV-2000.  
XX  
PF 11-MAY-2000; 2000WO-SE000930.  
XX  
PR 17-MAY-1999; 99SE-00001773.  
XX  
PA (STRO/) STROEMBERG N.  
PA (JOHA/) JOHANSSON I.  
XX  
PI Stroemberg N, Johansson I;  
XX  
DR WPI; 2001-031923/04.  
XX  
PT New oligopeptides comprising 2 arginine residues from degradation of  
PT proline-rich proteins, useful for preventing dental caries.  
XX  
PS Claim 4; Page 24; 36pp; English.  
XX  
CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
CC AAB48783) which contain at least two arginine residues and which protect  
CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
CC determinants of a person's susceptibility to dental caries. PRPs are  
CC degraded by Actinomyces and Streptococcus species to small peptide  
CC fragments. These are metabolised by oral bacteria for nutritional  
CC purposes, with certain bacterial species generating ammonia via the  
CC catabolism of arginine. The peptides of the invention, being arginine-  
CC rich, can also be converted to ammonia by these bacteria. The ammonia  
CC thus formed raises the pH at the dental surface, thereby protecting the  
CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-

CC derived oligopeptides of the invention

XX Sequence 5 AA;

Query Match 100.0%; Score 30; DB 4; Length 5;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
 |||||

Db 1 PPRGR 5

RESULT 2

AAV50279  
 ID AAY50279 standard; peptide; 6 AA.

XX AC AAY50279;

DT 12-JAN-2000 (first entry)

XX Neutrophil-activating pancreatic derived peptide 79.

XX Cell activation; pancreas; treatment; cardiovascular disease; trauma;  
 inflammatory disease; autoimmune diseases; arthritis; diabetes; stroke;  
 organ rejection; ischemia; Alzheimer's disease; myocardial infarction;  
 haemorrhagic shock; diabetic retinopathy; venous insufficiency; angina;  
 trauma; protease inhibitor; hypertension; sepsis.

XX Unidentified.

XX WO9946367-A2.

XX 16-SEP-1999.

XX 11-MAR-1999; 99WO-US005247.

XX 11-MAR-1998; 98US-00038894.

XX (CELL-) CELL ACTIVATION INC.

XX (REGC) UNIV CALIFORNIA.

XX (SCRI) SCRIPPS RES INST.

XX Stoughton RB, Schmid-Schonbein GW, Hugli TE, Kistler E;

XX WPI; 1999-580234/49.

XX Use of cell activating compositions in developing products for diagnosis  
 and treatment of e.g. cardiovascular, inflammatory, autoimmune or  
 Alzheimer's disease, trauma, arthritis, organ rejection, diabetes, stroke  
 or ischemia.

XX Example 9; Page 183; 184pp; English.

XX This invention describes a novel method for the use and preparation of  
 cell activating compositions which involves preparing a cell activating  
 composition comprising (a) homogenizing pancreatic tissue in buffer at  
 about neutral or higher pH to produce a homogenate; (b) removing  
 particulates from the homogenate; (c) optionally incubating the resulting  
 homogenate, with particulates removed, with a protease; and (d)  
 fractionating the homogenate and selecting fractions that exhibit cell  
 activation activity. The methods can be used for improving treatment  
 outcome or reducing risk of treatment of e.g. cardiovascular disease,  
 inflammatory disease, trauma, autoimmune diseases, arthritis, organ  
 rejection, diabetes and diabetic complications, stroke, ischemia,  
 Alzheimer's disease, myocardial infarction, haemorrhagic shock, diabetic  
 retinopathy, diabetes, venous insufficiency, unstable angina or trauma.  
 They can be used in the veterinary treatment of a non-human subject.  
 Protease inhibitors can be used to lower cell activation resulting from  
 these diseases and deficiencies. The detection of an elevated level of  
 hydrogen peroxide can be used to detect an inflammatory condition. An  
 elevated level of hydrogen peroxide in plasma or whole blood and in the  
 presence of superoxide dismutase (SOD) indicates leukocyte up regulation,

CC e.g. indicative of the onset of an acute cardiovascular disorders, such  
 as disease onset or ischemic complications. An elevated level of hydrogen  
 peroxide in plasma or whole blood and a low level in the presence of SOD  
 is indicative of a chronic or immune compromised condition e.g.  
 CC hypertension or sepsis. AAY50201-Y50334 represent peptides used in the  
 CC method of the invention

XX Sequence 6 AA;

Query Match 100.0%; Score 30; DB 2; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
 |||||

Db 2 PPRGR 6

RESULT 3

AAV48781  
 ID AAB48781 standard; peptide; 6 AA.

XX AC AAB48781;

DT 09-MAR-2001 (first entry)

XX Human saliva PRP-1 fragment (residues 103-108), SEQ ID NO:11.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;

XX Chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 oral bacterium; caries prevention.

XX Homo sapiens.

XX WO200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SE000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO/) STROEMBERG N.

XX (JOHA/) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.

XX Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 AAB48783) which contain at least two arginine residues and which protect  
 against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 encoded by six clustered genes on chromosome 12p13.2 and are potential  
 determinants of a person's susceptibility to dental caries. PRPs are  
 degraded by Actinomyces and Streptococcus species to small peptide  
 fragments. These are metabolised by oral bacteria for nutritional  
 purposes, with certain bacterial species generating ammonia via the  
 catabolism of arginine. The peptides of the invention, being arginine-  
 rich, can also be converted to ammonia by these bacteria. The ammonia  
 thus formed raises the pH at the dental surface, thereby protecting the  
 teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 derived oligopeptides of the invention

XX Sequence 6 AA;

Query Match 100.0%; Score 30; DB 4; Length 6;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

*difference by 2*

QY 1 PPRGR 5  
 DB 2 PPRGR 6

## RESULT 4

AA50278  
 ID AAY50278 standard; peptide; 7 AA.

AC AAY50278;

DT 12-JAN-2000 (first entry)

DE Neutrophil-activating pancreatic derived peptide 78.

XX Cell activation; pancreas; treatment; cardiovascular disease; trauma;  
 KW inflammatory disease; autoimmune diseases; arthritis; diabetes; stroke;  
 KW organ rejection; ischemia; Alzheimer's disease; myocardial infarction;  
 KW haemorrhagic shock; diabetic retinopathy; venous insufficiency; angina;  
 KW trauma; protease inhibitor; hypertension; sepsis.

XX Unidentified.

OS XX

XX XX

PN W09946367-A2.

XX 16-SEP-1999.

XX 11-MAR-1999; 99WO-US005247.

XX 11-MAR-1998; 98US-00038894.

XX (CELL-) CELL ACTIVATION INC.

PA (REGC) UNIV CALIFORNIA.

PA (SCRI) SCRIPPS RES INST.

XX Stoughton RB, Schmid-Schonbein GW, Hugli TE, Kistler E;  
 PI WPI; 1999-580234/49.

XX Use of cell activating compositions in developing products for diagnosis  
 PT and treatment of e.g. cardiovascular, inflammatory, autoimmune or  
 PT Alzheimer's disease, trauma, arthritis, organ rejection, diabetes, stroke  
 PT or ischemia.

XX Example 9; Page 183; 184pp; English.

XX This invention describes a novel method for the use and preparation of  
 CC cell activating compositions which involves preparing a cell activating  
 CC composition comprising (a) homogenizing pancreatic tissue in buffer at  
 CC about neutral or higher pH to produce a homogenate; (b) removing  
 CC particulates from the homogenate; (c) optionally incubating the resulting  
 CC homogenate, with particulates removed, with a protease; and (d)  
 CC fractionating the homogenate and selecting fractions that exhibit cell  
 CC activation activity. The methods can be used for improving treatment  
 CC outcome or reducing risk of treatment of e.g. cardiovascular disease,  
 CC inflammatory disease, trauma, autoimmune diseases, arthritis, organ  
 CC rejection, diabetes and diabetic complications, stroke, ischemia,  
 CC Alzheimer's disease, myocardial infarction, haemorrhagic shock, diabetic  
 CC retinopathy, diabetes, venous insufficiency, unstable angina or trauma.  
 CC They can be used in the veterinary treatment of a non-human subject.  
 CC Protease inhibitors can be used to lower cell activation resulting from  
 CC these diseases and deficiencies. The detection of an elevated level of  
 CC hydrogen peroxide can be used to detect an inflammatory condition. An  
 CC elevated level of hydrogen peroxide in plasma or whole blood and in the  
 CC presence of superoxide dismutase (SOD) indicates leukocyte up regulation,  
 CC e.g. indicative of the onset of an acute cardiovascular disorders, such  
 CC as disease onset or ischemic complications. An elevated level of hydrogen  
 CC peroxide in plasma or whole blood and a low level in the presence of SOD  
 CC is indicative of a chronic or immune compromised condition e.g.  
 CC hypertension or sepsis. AAY50201-Y50334 represent peptides used in the  
 CC method of the invention

SQ Sequence 7 AA;

Query Match 100.0%; Score 30; DB 2; Length 7;  
 Best Local Similarity 100.0%; Pred. NO. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
 DB 3 PPRGR 7

## RESULT 5

AA48780  
 ID AAB48780 standard; peptide; 7 AA.

XX AAB48780;

DT 09-MAR-2001 (first entry)

DE Human saliva PRP-1 fragment (residues 102-108), SEQ ID NO:10.

XX Human; PRP-1; proline-rich protein; saliva; dental caries;  
 KW Chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.

OS Homo sapiens.

PN W0200069890-A1.

XX 23-NOV-2000.

XX 11-MAY-2000; 2000WO-SE000930.

XX 17-MAY-1999; 99SE-00001773.

XX (STRO) STROEMBERG N.

PA (JOHA) JOHANSSON I.

XX Stroemberg N, Johansson I;

XX WPI; 2001-031923/04.

XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.

XX Claim 4; Page 24; 36pp; English.

XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-548783 represent the PRP-1-  
 CC derived oligopeptides of the invention

SQ Sequence 7 AA;

Query Match 100.0%; Score 30; DB 4; Length 7;  
 Best Local Similarity 100.0%; Pred. NO. 1.4e+06;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
 DB 3 PPRGR 7

## RESULT 6



PA (JOHA/) JOHANSSON I.  
 XX Stroenberg N, Johansson I;  
 PI WPI; 2001-031923/04.  
 XX  
 DR  
 PT New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 XX  
 FS Claim 4; Page 24; 36pp; English.  
 XX  
 XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention  
 XX  
 SQ Sequence 10 AA;  
 Query Match 100.0%; Score 30; DB 4; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 PPRGR 5  
 Db |||||  
 6 PPRGR 10  
 RESULT 9  
 ID AAB48783 standard; peptide; 17 AA.  
 XX AAB48783;  
 AC AAB48783;  
 XX  
 XX 09-MAR-2001 (first entry)  
 DT Human saliva PRP-1 fragment (residues 99-115), SEQ ID NO:13.  
 DE Human; PRP-1; proline-rich protein; saliva; dental caries;  
 KW chromosome 12p13.2; arginine catabolism; ammonia production; pH increase;  
 KW oral bacterium; caries prevention.  
 XX Homo sapiens.  
 OS WO200069890-A1.  
 XX  
 XX 23-NOV-2000.  
 PD 11-MAY-2000; 2000WO-SE000930.  
 PF 17-MAY-1999; 98SE-00001773.  
 XX (STRO/) STROENBERG N.  
 PA (JOHA/) JOHANSSON I.  
 XX Stroenberg N, Johansson I;  
 FI WPI; 2001-031923/04.  
 DR  
 XX New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 PT  
 XX Claim 2; Page 24; 36pp; English.  
 FS  
 XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention  
 XX  
 SQ Sequence 17 AA;  
 Query Match 100.0%; Score 30; DB 4; Length 17;  
 Best Local Similarity 100.0%; Pred. No. 91;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 PPRGR 5

PA (JOHA/) JOHANSSON I.  
 XX Stroenberg N, Johansson I;  
 PI WPI; 2001-031923/04.  
 XX  
 DR  
 PT New oligopeptides comprising 2 arginine residues from degradation of  
 PT proline-rich proteins, useful for preventing dental caries.  
 XX  
 FS Claim 4; Page 24; 36pp; English.  
 XX  
 XX The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
 CC AAB48783) which contain at least two arginine residues and which protect  
 CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
 CC encoded by six clustered genes on chromosome 12p13.2 and are potential  
 CC determinants of a person's susceptibility to dental caries. PRPs are  
 CC degraded by Actinomyces and Streptococcus species to small peptide  
 CC fragments. These are metabolised by oral bacteria for nutritional  
 CC purposes, with certain bacterial species generating ammonia via the  
 CC catabolism of arginine. The peptides of the invention, being arginine-  
 CC rich, can also be converted to ammonia by these bacteria. The ammonia  
 CC thus formed raises the pH at the dental surface, thereby protecting the  
 CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-  
 CC derived oligopeptides of the invention  
 XX  
 SQ Sequence 10 AA;  
 Query Match 100.0%; Score 30; DB 4; Length 10;  
 Best Local Similarity 100.0%; Pred. No. 57;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 PPRGR 5  
 Db |||||  
 6 PPRGR 10  
 RESULT 9  
 ID AAG93874 standard; peptide; 10 AA.  
 XX AAG93874;  
 AC AAG93874;  
 XX  
 XX 18-SEP-2001 (first entry)  
 DT Human complementary peptide, SEQ ID NO: 68.  
 DE Human; complementary peptide; ligand; drug discovery; drug design.  
 KW Homo sapiens.  
 OS WO200142277-A2.  
 XX  
 XX 14-JUN-2001.  
 PD 13-DEC-2000; 2000WO-GB004776.  
 PF 13-DEC-1999; 99GB-00029464.  
 XX (PROT-) PROTEOM LTD.  
 PA Roberts GW, Heal JR;  
 XX WPI; 2001-408419/43.  
 DR  
 XX A set of peptide ligands consisting of specific complementary peptides to  
 PT proteins encoded by genes of the human genome, useful in an assay for  
 PT screening and identifying of one or more novel peptides which are drug  
 PT candidates or pro-drugs.  
 XX  
 XX Example 4; Page 50; 645pp; English.  
 FS  
 XX The invention relates to a set of complementary peptide ligands generated  
 CC from the human genome. The complementary peptides interact with their

Db 6 PPRGR 10

|||||

## RESULT 11

AAU01685  
ID AAU01685 standard; protein, 25 AA.

XX AC AAU01685;

XX DT 18-JUL-2001 (first entry)

XX DE Gene 28 human secreted protein homologous amino acid sequence.

XX KW Human secreted protein; diagnosis; autoimmune disease;  
XX KW rheumatoid arthritis; hyperproliferative disorder; neoplasm; sunburn;  
XX KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;  
XX KW cerebral ischaemia; angioneuroma; nervous system disorder; skin aging;  
XX KW Alzheimer's disease; infection; ocular disorder; corneal infection;  
XX KW wound healing; epithelial cell proliferation; chemotaxis; preservative;  
XX KW organ transplantation; tissue regeneration; food additive.

XX OS Homo sapiens.

XX PN WO200123409-A2.

XX PD 05-APR-2001.

XX PF 26-SEP-2000; 2000WO-US026371.

XX PR 27-SEP-1999; 99US-0155804P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM, Komatsoulis GA;

XX DR WPI; 2001-266139/27.

XX PT Nucleic acids encoding 38 human secreted polypeptides, useful for  
XX PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's disease  
XX PT and diabetic retinopathy.

XX PS Disclosure; Page 47; 498pp; English.

XX CC AAU01641-AAU01698 represent human secreted protein amino acid, and  
XX CC related amino acid sequences of the invention. The human secreted protein  
XX CC sequences are used to prevent, treat or ameliorate a medical condition in  
XX CC e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep.  
XX CC They are also used in diagnosing a pathological condition or  
XX CC susceptibility to a pathological condition. The antibodies to human  
XX CC secreted proteins can also be used in alleviating symptoms associated  
XX CC with the disorders and in diagnostic immunoassays e.g. radioimmunoassays  
XX CC or enzyme linked immunosorbent assays (ELISA). Disorders which are  
XX CC diagnosed or treated include autoimmune diseases e.g. rheumatoid  
XX CC arthritis, hyperproliferative disorders e.g. neoplasms of the breast or  
XX CC liver, cardiovascular disorders e.g. cardiac arrest, cerebrovascular  
XX CC disorders e.g. cerebral ischaemia, angiogenesis, nervous system disorders  
XX CC e.g. Alzheimer's disease, infections caused by bacteria, viruses and  
XX CC fungi and ocular disorders e.g. corneal infection. The polypeptides can  
XX CC also be used to aid wound healing and epithelial cell proliferation, to  
XX CC prevent skin aging due to sunburn, to maintain organs before  
XX CC transplantation, for supporting cell culture of primary tissues, to  
XX CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
XX CC as a food additive or preservative to increase or decrease storage  
XX CC capabilities

XX SQ Sequence 25 AA;

## Query Match

Best Local Similarity 100.0%; Score 30; DB 4; Length 25;

Matches 5; Conservative 100.0%; Pred. No. 1.3e+02; Indels 0; Gaps 0;  
Mismatches 0;

QY 1 PPRGR 5

Db 18 PPRGR 22

|||||

## RESULT 12

AAU41484  
ID AAU41484 standard; protein, 28 AA.

XX AC AAU41484;

XX DT 02-DEC-1999 (first entry)

XX DE Fragment of human secreted protein encoded by gene 66.

XX KW Human; secreted protein; fusion protein; gene therapy; protein therapy;  
XX KW diagnosis; issue; cancer; tumour; neurodegenerative disorder; leukaemia;  
XX KW developmental abnormality; foetal deficiency; blood; allergy; renal;  
XX KW immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma;  
XX KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;  
XX KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;  
XX KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;  
XX KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

XX OS Homo sapiens.

XX PN WO9947540-A1.

XX PD 23-SEP-1999.

XX PF 18-MAR-1999; 99WO-US005804.

XX PR 19-MAR-1998; 98US-0078563P.

XX PR 19-MAR-1998; 98US-0078566P.

XX PR 19-MAR-1998; 98US-0078573P.

XX PR 19-MAR-1998; 98US-0078574P.

XX PR 19-MAR-1998; 98US-0078576P.

XX PR 19-MAR-1998; 98US-0078577P.

XX PR 19-MAR-1998; 98US-0078578P.

XX PR 19-MAR-1998; 98US-0078579P.

XX PR 01-APR-1998; 98US-0078581P.

XX PR 01-APR-1998; 98US-0080312P.

XX PR 01-APR-1998; 98US-0080313P.

XX PR 01-APR-1998; 98US-0080314P.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Ruben SM, Ni J, Rosen CA, Yu G, Young PE, Feng P, Soppet DR;  
XX PI Wei Y, Endress GA, Duan RD, Kyaw H, Ebner R, Lafleur DW, Olsen HS;  
XX PI Shi Y, Moore PA;

XX DR WPI; 1999-562050/47.

XX PT New isolated human genes, useful for diagnosis and treatment of e.g.  
XX PT cancers, neurological disorders, immune diseases, inflammation or blood  
XX PT disorders.

XX PS Disclosure; Page 118; 484pp; English.

XX CC This sequence represents a fragment of a secreted human protein encoded  
XX CC by the nucleic acid molecule detailed in the descriptor line. The gene  
XX CC can be used to generate fusion proteins by linking to the gene to a human  
XX CC immunoglobulin Fc portion (e.g. AAU24802) for increasing the stability of  
XX CC the fused protein as compared to the human protein only. The invention  
XX CC relates to 95 novel genes and their fragments (nucleic acid sequences:  
XX CC AAU24811-224907; amino acid sequences AAU1308-Y41404) which are useful  
XX CC for preventing, treating or ameliorating medical conditions e.g. by  
XX CC protein or gene therapy. Also, pathological conditions can be diagnosed  
XX CC by determining the amount of the new polypeptides in a sample or by  
XX CC determining the presence of mutations in the new polynucleotides.  
XX CC Specific uses are described for each of the 95 polynucleotides based on  
XX CC which tissues they are most highly expressed in (see AAU24811 for  
XX CC described uses)

## SQ Sequence 28 AA;

Query Match 100.0%; Score 30; DB 2; Length 28;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
 |||||  
 Db 2 PPRGR 6

## RESULT 13

AAAY00630  
 ID AAY00630 standard; peptide; 34 AA.

XX AC AAY00630;

DT 26-JUL-1999 (first entry)

XX Human telomerase protein sequence fragment.

XX Telomerase; human; cancer; diagnosis; melanoma; skin cancer; leukaemia;  
 KW neuroblastoma; breast carcinoma; colon carcinoma; lymphoma; osteosarcoma;  
 KW smooth muscle cell hyperplasia; stem cell proliferation; Wilm's tumour;  
 KW stem cell differentiation; organ regeneration; organ differentiation.

XX Homo sapiens.

XX WO9901560-A1.

PD 14-JAN-1999.

XX 01-JUL-1998; 98WO-US013835.

XX 01-JUL-1997; 97US-0051410P.

XX 21-JUL-1997; 97US-0053018P.

XX 21-JUL-1997; 97US-0053229P.

XX 04-AUG-1997; 97US-0054642P.

XX 09-SEP-1997; 97US-0058287P.

XX (CAMB-) CAMBIA BIOSYSTEMS LLC.

XX Kilian A, Bowtell D;

XX WPI; 1999-106060/09.

XX N-PSDB; AAX18255.

XX Claim 6; Fig 10a; 134pp; English.

XX This sequence is a fragment of the human telomerase of the invention.  
 CC Primers that amplify the telomerase coding sequence can be used in a  
 CC method for diagnosing cancer in a patient. The telomerase can be used for  
 CC detection, diagnosis and drug screening. Inhibitors of telomerase  
 CC activity can be used to treat cancers such as melanomas, other skin  
 CC cancers, neuroblastomas, breast carcinomas, colon carcinomas, leukaemias,  
 CC lymphomas, osteosarcomas or smooth muscle cell hyperplasias or skin  
 CC growths. Enhancers of telomerase may be used to stimulate stem cell  
 CC proliferation and differentiation (expansion of haematopoietic stem cells  
 CC could be administered in the bone marrow transplant context). As well,  
 CC many tissues have stem cells. Proliferation of these cells may be useful  
 CC in wound healing, hair growth, treatment of disease such as Wilm's  
 CC tumour, organ regeneration or differentiation after injury or diseases,  
 CC nerve cell or brain cell growth following injury

XX Sequence 34 AA;

Query Match 100.0%; Score 30; DB 2; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
 |||||  
 Db 1 PPRGR 5

## RESULT 14

AAAY00658  
 ID AAY00658 standard; protein; 34 AA.

XX AC AAY00658;

DT 26-JUL-1999 (first entry)

XX Telomerase protein sequence fragment.

XX Telomerase; human; cancer; diagnosis; melanoma; skin cancer; leukaemia;  
 KW neuroblastoma; breast carcinoma; colon carcinoma; lymphoma; osteosarcoma;  
 KW smooth muscle cell hyperplasia; stem cell proliferation; Wilm's tumour;  
 KW stem cell differentiation; organ regeneration; organ differentiation; ss.

XX Homo sapiens.

XX Synthetic.

XX WO9901560-A1.

XX 14-JAN-1999.

XX 01-JUL-1998; 98WO-US013835.

XX 01-JUL-1997; 97US-0051410P.

XX 21-JUL-1997; 97US-0053018P.

XX 21-JUL-1997; 97US-0053229P.

XX 04-AUG-1997; 97US-0054642P.

XX 09-SEP-1997; 97US-0058287P.

XX (CAMB-) CAMBIA BIOSYSTEMS LLC.

XX Kilian A, Bowtell D;

XX WPI; 1999-106060/09.

XX N-PSDB; AAX18284.

XX Claim 4; Fig 11; 134pp; English.

XX This sequence encodes a fragment of a truncated human telomerase of the  
 CC invention. Primers that amplify the telomerase coding sequence can be  
 CC used in a method for diagnosing cancer in a patient. The telomerase can  
 CC be used for detection, diagnosis and drug screening. Inhibitors of  
 CC telomerase activity can be used to treat cancers such as melanomas, other  
 CC skin cancers, neuroblastomas, breast carcinomas, colon carcinomas, other  
 CC leukaemias, lymphomas, osteosarcomas or smooth muscle cell hyperplasias  
 CC or skin growths. Enhancers of telomerase may be used to stimulate stem  
 CC cell proliferation and differentiation (expansion of haematopoietic stem  
 CC cells could be administered in the bone marrow transplant context). As  
 CC well, many tissues have stem cells. Proliferation of these cells may be  
 CC useful in wound healing, hair growth, treatment of disease such as Wilm's  
 CC tumour, organ regeneration or differentiation after injury or diseases,  
 CC nerve cell or brain cell growth following injury. Note: this sequence is  
 CC used to replace the N-terminus of some of the full length human  
 CC telomerase protein sequences of the invention

XX Sequence 34 AA;

Query Match 100.0%; Score 30; DB 2; Length 34;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5



Search completed: April 6, 2004, 16:06:42  
Job time : 28.9907 secs

Db 1 PPRGR 5  
|||||  
RESULT 15  
ABP62124  
ID ABP62124 standard; protein; 44 AA.  
XX AC ABP62124;  
XX DT 12-NOV-2002 (first entry)  
XX DE Human secreted protein SEQ ID NO 177.  
XX KW Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;  
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnary;  
KW antiparkinsonian; antiskinkling; antianaemic; antiarthritis; cancer;  
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
KW antiallergic; antidiabetic; antifungal; anticonvulsant; antifungal;  
KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; nephrotropic; gene therapy; vaccine.  
XX OS Homo sapiens.  
XX PN WO200257420-A2.  
XX PD 25-JUL-2002.  
XX PF 17-JAN-2002; 2002WO-US001109.  
XX PR 18-JAN-2001; 2001US-0262066P.  
XX PA (HUMA-) HUMAN GENOME SCI INC.  
XX PI Moore PA, Ruben SM, Lafleur DW, Shi Y, Rosen CA, Olsen H;  
PI Ebner R, Brewer LA;  
XX WPI 2002-599716/64.  
XX New polynucleotides and polypeptides useful for diagnosing, prognosing,  
PT treating or preventing e.g. neurodegenerative, central nervous system,  
PT autoimmune, respiratory, reproductive, or inflammatory diseases or  
PT disorders.  
XX Claim 11; Page 54; 785pp; English.  
XX The invention relates to novel genes (ABQ92553-ABQ92607) and proteins  
CC (ABP62013-ABP62153) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)  
CC cardiovascular disorders such as myocardial ischaemias; (d) wound healing  
CC ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
CC infectious diseases such as viral, bacterial, fungal and parasitic  
XX infections  
SQ Sequence 44 AA;  
Query Match 100.0%; Score 30; DB 5; Length 44;  
Best Local Similarity 100.0%; Pred.No. 2.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 PPRGR 5  
|||||  
Db 6 PPRGR 10

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 16:14:50 ; Search time 19.6262 Seconds  
(without alignments)  
66.909 Million cell updates/sec

Title: US-10-009-709-12

Perfect score: 30

Sequence: 1 PPRGR 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 26263353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PTC\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PTCUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	30	100.0	10	10	US-09-572-404B-68
2	30	100.0	12	9	US-09-908-322-48
3	30	100.0	12	10	US-09-783-931-48
4	30	100.0	28	12	US-10-653-585-302
5	30	100.0	40	12	US-10-424-599-235436
6	30	100.0	40	14	US-10-029-386-28348
7	30	100.0	44	14	US-10-411-224-177
8	30	100.0	44	15	US-10-047-021-177
9	30	100.0	46	9	US-09-864-761-40999
10	30	100.0	51	12	US-10-424-599-172569
11	30	100.0	63	12	US-10-424-599-238344
12	30	100.0	65	12	US-10-424-599-277753
13	30	100.0	72	12	US-10-424-599-231007
14	30	100.0	67	12	US-10-424-599-204287
15	30	100.0	72	12	US-10-424-599-223703

16	30	100.0	79	12	US-10-424-599-236933
17	30	100.0	80	12	US-10-424-599-232682
18	30	100.0	88	9	US-09-764-864-1286
19	30	100.0	89	12	US-10-424-599-229053
20	30	100.0	89	12	US-10-424-599-279722
21	30	100.0	92	12	US-10-424-599-178114
22	30	100.0	94	12	US-10-424-599-285441
23	30	100.0	96	12	US-10-424-599-191513
24	30	100.0	101	12	US-10-424-599-185963
25	30	100.0	103	12	US-10-425-114-40547
26	30	100.0	105	12	US-10-424-599-242491
27	30	100.0	105	14	US-10-156-761-11675
28	30	100.0	108	12	US-10-424-599-142896
29	30	100.0	110	12	US-10-425-114-57743
30	30	100.0	111	12	US-10-424-599-163482
31	30	100.0	111	12	US-10-424-599-177430
32	30	100.0	112	12	US-10-424-599-269841
33	30	100.0	118	12	US-10-424-599-242444
34	30	100.0	122	12	US-10-424-599-197307
35	30	100.0	126	12	US-10-424-599-239125
36	30	100.0	129	12	US-10-425-114-51211
37	30	100.0	132	9	US-09-864-761-43644
38	30	100.0	138	12	US-10-424-599-276206
39	30	100.0	139	15	US-10-264-049-3161
40	30	100.0	152	12	US-10-424-599-210951
41	30	100.0	154	9	US-09-864-761-41874
42	30	100.0	161	9	US-09-821-687-2
43	30	100.0	161	14	US-10-017-161-1070
44	30	100.0	161	15	US-10-292-798-908
45	30	100.0	163	12	US-10-425-114-66926

#### ALIGNMENTS

##### RESULT 1

US-09-572-404B-68  
; Sequence 68, Application US/09572404B  
; Publication No. US20030078374A1  
; GENERAL INFORMATION:  
; APPLICANT: Proteom Ltd  
; TITLE OF INVENTION: Complementary peptide ligands from the human genome  
; FILE REFERENCE: Human patent  
; CURRENT APPLICATION NUMBER: US/09/572,404B  
; CURRENT FILING DATE: 2000-05-17  
; NUMBER OF SEQ ID NOS: 4203  
; SOFTWARE: ProtPatent version 1.0  
; SEQ ID NO 68  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: Homo Sapiens  
; FEATURE:  
; OTHER INFORMATION: sequence located in PIK3CA at 24-33 and may interact with Sequ  
US-09-572-404B-68

Query Match 100.0%; Score 30; DB 10; Length 10;  
Best Local Similarity 100.0%; Pred. No. 71;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5

Db 6 PPRGR 10

##### RESULT 2

US-09-908-322-48  
; Sequence 48, Application US/09908322  
; Patent No. US20020107194A1  
; GENERAL INFORMATION:  
; APPLICANT: Ish-Horowitz, David  
; Henrique, Domingos Manuel Pinto  
; Lewis, Julian Hart

```
;
;
; Artavanis-Tsakonas, Spyridon
; Gray, Grace
;
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF
; VERTEBRATE DELTA GENE AND METHODS BASED THEREON
;
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
;
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/908,322
; FILING DATE: 17-Jul-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/981,392
; FILING DATE: 22-DEC-1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-123
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
;
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:
;
; US-09-908-322-48
;
; Query Match 100.0%; Score 30; DB 9; Length 12;
; Best Local Similarity 100.0%; Pred.No. 84;
; Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 PPRGR 5
; DB 5 PPRGR 9
;
; RESULT 3
; US-09-783-931-48
; Sequence 48, Application US/09783931
; Publication No. US20030073620A1
;
; GENERAL INFORMATION:
; APPLICANT: Ish-Horowitz, David
; Henrique, Domingos Manuel Pinto
; Lewis, Julian Hart
; Artavanis-Tsakonas, Spyridon
; Gray, Grace
;
; TITLE OF INVENTION: ANTIBODIES TO VERTEBRATE DELTA PROTEINS
;
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036/2711
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
```

```
;
;
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/783,931
; FILING DATE: 15-Feb-2001
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/981,392
; FILING DATE: 22-DEC-1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Antler, Adriane M.
; REGISTRATION NUMBER: 32,605
; REFERENCE/DOCKET NUMBER: 7326-122
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-790-9090
; TELEFAX: 212-869-8864
; TELEX: 66141 PENNIE
;
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 48:
;
; US-09-783-931-48
;
; Query Match 100.0%; Score 30; DB 10; Length 12;
; Best Local Similarity 100.0%; Pred.No. 84;
; Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 1 PPRGR 5
; DB 5 PPRGR 9
;
; RESULT 4
; US-10-653-595-302
; Sequence 302, Application US/10653595
; Publication No. US20040048304A1
;
; GENERAL INFORMATION:
; APPLICANT: Ruben et. al.
; TITLE OF INVENTION: 95 Human secreted proteins
; FILE REFERENCE: P2027P1C1
; CURRENT APPLICATION NUMBER: US/10/653,595
; CURRENT FILING DATE: 2003-09-03
; PRIOR APPLICATION NUMBER: US 09/397945
; PRIOR FILING DATE: 1999-09-17
; PRIOR APPLICATION NUMBER: PCT/US99/05804
; PRIOR FILING DATE: 1999-03-18
; PRIOR APPLICATION NUMBER: 60/078,566
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,576
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,573
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,574
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/078,579
; PRIOR FILING DATE: 1998-03-19
; PRIOR APPLICATION NUMBER: 60/080,314
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/080,312
; PRIOR FILING DATE: 1998-04-01
; PRIOR APPLICATION NUMBER: 60/078,578
; PRIOR FILING DATE: 1998-03-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 470
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 302
; LENGTH: 28
; TYPE: PRT
```

ORGANISM: Homo sapiens  
US-10-653-595-302

Query Match 100.0%; Score 30; DB 12; Length 28;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
Db 2 PPRGR 6

## RESULT 5

US-10-424-599-235436  
; Sequence 235436, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 235436  
; LENGTH: 40  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_54626C.1.pep  
US-10-424-599-235436

Query Match 100.0%; Score 30; DB 12; Length 40;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
Db 11 PPRGR 15

## RESULT 6

US-10-029-386-28348  
; Sequence 28348, Application US/10029386  
; Publication No. US20030194704A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO  
; FILE REFERENCE: AEOICA-X-2  
; CURRENT APPLICATION NUMBER: US/10/029,386  
; CURRENT FILING DATE: 2001-12-20  
; NUMBER OF SEQ ID NOS: 34288  
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
; SEQ ID NO 28348  
; LENGTH: 40  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: MAP TO CHR22.125.0  
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.33  
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.49  
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.48  
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.39  
US-10-029-386-28348

Query Match 100.0%; Score 30; DB 14; Length 40;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 PPRGR 5  
Db 24 PPRGR 28

## RESULT 7

US-10-411-224-177  
; Sequence 177, Application US/10411224  
; Publication No. US2003016906A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 50 Human Secreted Proteins  
; FILE REFERENCE: P2016P1  
; CURRENT APPLICATION NUMBER: US/10/411,224  
; CURRENT FILING DATE: 2003-04-11  
; PRIOR APPLICATION NUMBER: US/09/722,329  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: 09/262,109  
; PRIOR FILING DATE: 1998-03-04  
; PRIOR APPLICATION NUMBER: 60/057,626  
; PRIOR FILING DATE: 1997-09-05  
; PRIOR APPLICATION NUMBER: 60/057,663  
; PRIOR FILING DATE: 1997-09-05  
; PRIOR APPLICATION NUMBER: 60/057,669  
; PRIOR FILING DATE: 1997-09-05  
; PRIOR APPLICATION NUMBER: 60/058,667  
; PRIOR FILING DATE: 1997-09-12  
; PRIOR APPLICATION NUMBER: 60/058,974  
; PRIOR FILING DATE: 1997-09-12  
; PRIOR APPLICATION NUMBER: 60/058,973  
; PRIOR FILING DATE: 1997-09-12  
; PRIOR APPLICATION NUMBER: 60/058,666  
; PRIOR FILING DATE: 1997-09-12  
; PRIOR APPLICATION NUMBER: 60/090,112  
; PRIOR FILING DATE: 1998-06-22  
; NUMBER OF SEQ ID NOS: 206  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 177  
; LENGTH: 44  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-411-224-177

Query Match 100.0%; Score 30; DB 14; Length 44;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
Db 6 PPRGR 10

## RESULT 8

US-10-047-021-177  
; Sequence 177, Application US/10047021  
; Publication No. US20040002591A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 50 Human Secreted Proteins  
; FILE REFERENCE: P2016P2  
; CURRENT APPLICATION NUMBER: US/10/047,021  
; CURRENT FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: US 60/262,066  
; PRIOR FILING DATE: 2001-01-18  
; PRIOR APPLICATION NUMBER: US 09/722,329  
; PRIOR FILING DATE: 2000-11-28  
; PRIOR APPLICATION NUMBER: US 09/262,109  
; PRIOR FILING DATE: 1999-03-04  
; PRIOR APPLICATION NUMBER: PCT/US98/18360  
; PRIOR FILING DATE: 1998-09-03  
; PRIOR APPLICATION NUMBER: US 60/057,626

PRIOR FILING DATE: 1997-09-05  
PRIOR APPLICATION NUMBER: US 60/057,663  
PRIOR FILING DATE: 1997-09-05  
PRIOR APPLICATION NUMBER: US 60/057,669  
PRIOR FILING DATE: 1997-09-05  
PRIOR APPLICATION NUMBER: US 60/058,667  
PRIOR FILING DATE: 1997-09-12  
PRIOR APPLICATION NUMBER: US 60/058,974  
PRIOR FILING DATE: 1997-09-12  
PRIOR APPLICATION NUMBER: US 60/058,973  
PRIOR FILING DATE: 1997-09-12  
PRIOR APPLICATION NUMBER: US 60/058,666  
PRIOR FILING DATE: 1997-09-12  
PRIOR APPLICATION NUMBER: US 60/090,112  
PRIOR FILING DATE: 1998-06-22  
NUMBER OF SEQ ID NOS: 206  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 177  
LENGTH: 44  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-047-021-177

Query Match 100.0%; Score 30; DB 15; Length 44;  
Best Local Similarity 100.0%; Pred. No. 2.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
|||  
Db 6 PPRGR 10

RESULT 9  
US-09-864-761-40999  
Sequence 40999, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aeonica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GS 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661

PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 40999  
LENGTH: 46  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL136365.3  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.9  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5  
OTHER INFORMATION: SWISSPROT HIT: P30516, EVALUATE 4.80e+00  
OTHER INFORMATION: EST\_HUMAN HIT: BE250573.1, EVALUATE 4.00e+00  
US-09-864-761-40999

Query Match 100.0%; Score 30; DB 9; Length 46;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
|||  
Db 11 PPRGR 15

RESULT 10  
US-10-424-599-172569  
Sequence 172569, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yinhua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223) B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 172569  
LENGTH: 51  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_126846C.1.ppt  
US-10-424-599-172569

Query Match 100.0%; Score 30; DB 12; Length 51;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
|||  
Db 40 PPRGR 44

RESULT 11  
US-10-424-599-238344  
Sequence 238344, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 238344  
; LENGTH: 63  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_5724C.1.pap  
US-10-424-599-238344

Query Match 100.0%; Score 30; DB 12; Length 63;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
Db 27 PPRGR 31

RESULT 12  
US-10-424-599-277753  
; Sequence 277753, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 277753  
; LENGTH: 65  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_92834C.1.pap  
US-10-424-599-277753

Query Match 100.0%; Score 30; DB 12; Length 65;  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
Db 41 PPRGR 45

RESULT 13  
US-10-424-599-231007  
; Sequence 231007, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 231007  
; LENGTH: 67  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_50620C.1.pap  
US-10-424-599-231007

Query Match 100.0%; Score 30; DB 12; Length 67;  
Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
Db 52 PPRGR 56

RESULT 14  
US-10-424-599-204287  
; Sequence 204287, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 204287  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_26499C.1.pap  
US-10-424-599-204287

Query Match 100.0%; Score 30; DB 12; Length 72;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
Db 22 PPRGR 26

RESULT 15  
US-10-424-599-223703  
; Sequence 223703, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 223703  
; LENGTH: 72  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_44033C.1.pap  
US-10-424-599-223703

Tue Apr 6 17:16:03 2004

Query Match 100.0%; Score 30; DB 12; Length 72;  
Best Local Similarity 100.0%; Pred. No. 4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 PPRGR 5  
Db 29 PPRGR 33

Search completed: April 6, 2004, 17:06:09  
Job time : 20.6262 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:56:34 ; Search time 7.3645 Seconds  
(without alignments)  
35.185 Million cell updates/sec

Title: US-10-009-709-12

Perfect score: 30

Sequence: 1 PPRGR 5

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/prodata/2/iaa/5A-COMB.pep.\*  
2: /cgn2\_6/prodata/2/iaa/5B-COMB.pep.\*  
3: /cgn2\_6/prodata/2/iaa/5A-COMB.pep.\*  
4: /cgn2\_6/prodata/2/iaa/5B-COMB.pep.\*  
5: /cgn2\_6/prodata/2/iaa/5A-COMB.pep.\*  
6: /cgn2\_6/prodata/2/iaa/5B-COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	30	100.0	12	US-08-981-392-48	Sequence 48, Appl
2	30	100.0	81	US-09-252-991A-30895	Sequence 30895, A
3	30	100.0	98	US-08-150-203A-4	Sequence 4, Appl
4	30	100.0	98	US-08-454-730-4	Sequence 4, Appl
5	30	100.0	98	US-08-949-788-4	Sequence 4, Appl
6	30	100.0	111	US-09-252-991A-23343	Sequence 23343, A
7	30	100.0	139	US-09-252-991A-22741	Sequence 22741, A
8	30	100.0	142	US-09-252-991A-23673	Sequence 23673, A
9	30	100.0	157	US-09-252-991A-31989	Sequence 31989, A
10	30	100.0	159	US-09-252-991A-31988	Sequence 31988, A
11	30	100.0	165	US-09-252-991A-21773	Sequence 21773, A
12	30	100.0	172	US-09-252-991A-18283	Sequence 18283, A
13	30	100.0	172	US-09-252-991A-18419	Sequence 18419, A
14	30	100.0	175	US-09-252-991A-28806	Sequence 28806, A
15	30	100.0	177	US-09-252-991A-29848	Sequence 29848, A
16	30	100.0	180	US-08-483-533-29	Sequence 29, Appl
17	30	100.0	180	US-09-283-471A-29	Sequence 29, Appl
18	30	100.0	183	US-09-252-991A-21850	Sequence 21850, A
19	30	100.0	183	US-09-252-991A-24990	Sequence 24990, A
20	30	100.0	195	US-09-252-991A-21451	Sequence 21451, A
21	30	100.0	196	US-09-252-991A-23646	Sequence 23646, A
22	30	100.0	207	US-08-559-397A-13	Sequence 13, Appl
23	30	100.0	207	US-09-252-991A-29505	Sequence 29505, A
24	30	100.0	209	US-08-559-397A-11	Sequence 11, Appl
25	30	100.0	223	US-09-252-991A-17072	Sequence 17072, A
26	30	100.0	225	US-09-252-991A-27105	Sequence 27105, A
27	30	100.0	228	US-09-436-983-7	Sequence 7, Appl

28 30 100.0 234 3 US-09-436-983-6 Sequence 6, Appli  
29 30 100.0 238 6 5223425-5 Patent No. 5223425  
30 30 100.0 239 4 US-09-252-991A-21942 Sequence 21942, A  
31 30 100.0 242 4 US-09-252-991A-25956 Sequence 25956, A  
32 30 100.0 250 6 5223425-4 Patent No. 5223425  
33 30 100.0 259 4 US-09-252-991A-28573 Sequence 28573, A  
34 30 100.0 263 4 US-09-489-039A-14069 Sequence 14069, A  
35 30 100.0 264 4 US-09-252-991A-16625 Sequence 16625, A  
36 30 100.0 296 4 US-09-252-991A-27429 Sequence 27429, A  
37 30 100.0 296 4 US-09-252-991A-32477 Sequence 32477, A  
38 30 100.0 302 4 US-09-252-991A-19798 Sequence 19798, A  
39 30 100.0 305 4 US-09-252-991A-25333 Sequence 25333, A  
40 30 100.0 307 4 US-09-134-000C-5027 Sequence 5027, Ap  
41 30 100.0 319 4 US-09-252-991A-23566 Sequence 23566, A  
42 30 100.0 325 4 US-09-252-991A-21743 Sequence 21743, A  
43 30 100.0 334 4 US-09-252-991A-22532 Sequence 22532, A  
44 30 100.0 350 6 5352575-7 Patent No. 5352575  
45 30 100.0 355 3 US-08-483-533-41 Sequence 41, Appl

#### ALIGNMENTS

RESULT 1  
US-08-981-392-48  
; Sequence 48, Application US/08981392  
; Patent No. 6262025  
; GENERAL INFORMATION:  
; APPLICANT: Ish-Horowitz, David  
; APPLICANT: Henrique, Domingos Manuel Pinto  
; APPLICANT: Lewis, Julian Hart  
; APPLICANT: Attavanis-Teakonas, Spyridon  
; APPLICANT: Gray, Grace  
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES  
; TITLE OF INVENTION: OF VERTEBRATE DELTA GENES AND METHODS BASED THEREON  
; NUMBER OF SEQUENCES: 94  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds LLP  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10036/2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/981,392  
; FILING DATE: 22-DEC-1997  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Antler, Adriane M.  
; REGISTRATION NUMBER: 32,605  
; REFERENCE/DOCKET NUMBER: 7326-038  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-790-9090  
; TELEFAX: 212-869-8864  
; INFORMATION FOR SEQ ID NO: 48:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 12 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
; US-08-981-392-48

Query Match 100.0%; Score 30; DB 3; Length 12;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;



QY 1 PPRGR 5  
|  
|  
|  
|  
|  
Db 5 PPRGR 9

## RESULT 2

US-09-252-991A-30895  
; Sequence 30895, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252.991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074.788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094.190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 30895  
; LENGTH: 81  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-30895

Query Match 100.0%; Score 30; DB 4; Length 81;  
Best Local Similarity 100.0%; Pred. No. 68;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
|  
|  
|  
|  
|  
Db 69 PPRGR 73

## RESULT 3

US-08-150-203A-4  
; Sequence 4, Application US/08150203A  
; Patent No. 5676951  
; GENERAL INFORMATION:  
; APPLICANT: Rijsewijk, Franciscus Antonius Maria  
; APPLICANT: van Oirschot, Johannes Theodorus  
; APPLICANT: Maes, Roger Kamiel  
; TITLE OF INVENTION: Bovine Herpesvirus Type 1  
; TITLE OF INVENTION: Deletion Mutants, Vaccines Based  
; TITLE OF INVENTION: Theron, Diagnostic Kits For  
; TITLE OF INVENTION: Detection Of Bovine Herpesvirus  
; TITLE OF INVENTION: Type 1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hoffmann & Baron  
; STREET: 350 Jericho Turnpike  
; CITY: Jericho  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 11753  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette - 3.5 inch,  
; MEDIUM TYPE: 1.44 MB Storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS DOS  
; SOFTWARE: WORD PERFECT 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/150.203A  
; FILING DATE: December 6, 1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Louise A. Fouch  
; REGISTRATION NUMBER: 37,133  
; REFERENCE/DOCKET NUMBER: 294-22  
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (516) 822-3550  
; TELEFAX: (516) 822-3582  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 98 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: Linear  
US-08-150-203A-4

Query Match 100.0%; Score 30; DB 1; Length 98;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5  
|  
|  
|  
|  
|  
Db 92 PPRGR 96

## RESULT 4

US-08-454-730-4  
; Sequence 4, Application US/08454730  
; Patent No. 5789177  
; GENERAL INFORMATION:  
; APPLICANT: Rijsewijk, Franciscus Antonius Maria  
; APPLICANT: van Oirschot, Johannes Theodorus  
; APPLICANT: Maes, Roger Kamiel  
; TITLE OF INVENTION: Bovine Herpesvirus Type 1  
; TITLE OF INVENTION: Deletion Mutants, Vaccines Based  
; TITLE OF INVENTION: Theron, Diagnostic Kits For  
; TITLE OF INVENTION: Detection Of Bovine Herpesvirus  
; TITLE OF INVENTION: Type 1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hoffmann & Baron  
; STREET: 350 Jericho Turnpike  
; CITY: Jericho  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 11753  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette - 3.5 inch,  
; MEDIUM TYPE: 1.44 MB Storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS DOS  
; SOFTWARE: WORD PERFECT 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/454.730  
; FILING DATE: May 31, 1995  
; CLASSIFICATION: 536  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/150.203  
; FILING DATE: December 6, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ronald J. Baron  
; REGISTRATION NUMBER: 29,281  
; REFERENCE/DOCKET NUMBER: 294-22 DIV  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (516) 822-3550  
; TELEFAX: (516) 822-3582  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 98 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: Linear  
US-08-454-730-4

Query Match 100.0%; Score 30; DB 1; Length 98;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5

Db 92 PPRGR 96

RESULT 5  
US-08-949-788-4  
; Sequence 4, Application US/09949788  
; Patent No. 6403097  
; GENERAL INFORMATION:  
; APPLICANT: Rijsewijk, Franciscus Antonius Maria  
; APPLICANT: van Oirschot, Johannes Theodorus  
; APPLICANT: Maes, Roger Kamel  
; TITLE OF INVENTION: Bovine Herpesvirus Type 1  
; TITLE OF INVENTION: Deletion Mutants, Vaccines Based  
; TITLE OF INVENTION: Theorem, Diagnostic Kits For  
; TITLE OF INVENTION: Detection Of Bovine Herpesvirus  
; TITLE OF INVENTION: Type 1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hoffmann & Baron  
; STREET: 350 Jericho Turnpike  
; CITY: Jericho  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 11753  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette - 3.5 inch,  
; MEDIUM TYPE: 1.44 MB Storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS DOS  
; SOFTWARE: WORD PERFECT 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/949,788  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/150,203  
; FILING DATE: 22-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Louise A. Fouch  
; REGISTRATION NUMBER: 37,133  
; REFERENCE/DOCKET NUMBER: 294-22  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (516) 822-3550  
; TELEFAX: (516) 822-3582  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 98 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: Linear  
US-08-949-788-4

Query Match 100.0%; Score 30; DB 4; Length 98;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
Db 92 PPRGR 96

RESULT 6  
US-09-252-991A-23343  
; Sequence 23343, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18

Query Match 100.0%; Score 30; DB 4; Length 111;  
Best Local Similarity 100.0%; Pred. No. 92;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
Db 74 PPRGR 78

RESULT 7  
US-09-252-991A-22741  
; Sequence 22741, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 22741  
; LENGTH: 139  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-22741

Query Match 100.0%; Score 30; DB 4; Length 139;  
Best Local Similarity 100.0%; Pred. No. 1.1e-02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
Db 21 PPRGR 25

RESULT 8  
US-09-252-991A-23673  
; Sequence 23673, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23673  
; LENGTH: 142  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23673

Db 92 PPRGR 96

RESULT 5  
US-08-949-788-4  
; Sequence 4, Application US/09949788  
; Patent No. 6403097  
; GENERAL INFORMATION:  
; APPLICANT: Rijsewijk, Franciscus Antonius Maria  
; APPLICANT: van Oirschot, Johannes Theodorus  
; APPLICANT: Maes, Roger Kamel  
; TITLE OF INVENTION: Bovine Herpesvirus Type 1  
; TITLE OF INVENTION: Deletion Mutants, Vaccines Based  
; TITLE OF INVENTION: Theorem, Diagnostic Kits For  
; TITLE OF INVENTION: Detection Of Bovine Herpesvirus  
; TITLE OF INVENTION: Type 1  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hoffmann & Baron  
; STREET: 350 Jericho Turnpike  
; CITY: Jericho  
; STATE: New York  
; COUNTRY: United States of America  
; ZIP: 11753  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette - 3.5 inch,  
; MEDIUM TYPE: 1.44 MB Storage  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS DOS  
; SOFTWARE: WORD PERFECT 6.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/949,788  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/150,203  
; FILING DATE: 22-MAR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Louise A. Fouch  
; REGISTRATION NUMBER: 37,133  
; REFERENCE/DOCKET NUMBER: 294-22  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (516) 822-3550  
; TELEFAX: (516) 822-3582  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 98 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: Linear  
US-08-949-788-4

Query Match 100.0%; Score 30; DB 4; Length 98;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
Db 92 PPRGR 96

RESULT 6  
US-09-252-991A-23343  
; Sequence 23343, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18

```
Query Match      100.0%; Score 30; DB 4; Length 142;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5
    |||||
Db 60 PPRGR 64

RESULT 9
US-09-252-991A-31989
; Sequence 31989, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31989
; LENGTH: 157
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31989

Query Match      100.0%; Score 30; DB 4; Length 157;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5
    |||||
Db 71 PPRGR 75

RESULT 10
US-09-252-991A-31988
; Sequence 31988, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31988
; LENGTH: 159
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31988

Query Match      100.0%; Score 30; DB 4; Length 159;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5
    |||||
Db 41 PPRGR 45

RESULT 11
US-09-252-991A-21773
; Sequence 21773, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21773
; LENGTH: 165
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21773

Query Match      100.0%; Score 30; DB 4; Length 165;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5
    |||||
Db 149 PPRGR 153

RESULT 12
US-09-252-991A-18283
; Sequence 18283, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18283
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18283

Query Match      100.0%; Score 30; DB 4; Length 172;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5
    |||||
Db 137 PPRGR 141

RESULT 13
US-09-252-991A-18419
; Sequence 18419, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18419
; LENGTH: 172
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18419

Query Match      100.0%; Score 30; DB 4; Length 172;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PPRGR 5
    |||||
Db 137 PPRGR 141
```

; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 18419  
; LENGTH: 172  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-18419

Query Match 100.0%; Score 30; DB 4; Length 172;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
Db 113 PPRGR 117

## RESULT 14

US-09-252-991A-28806  
; Sequence 28806, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 28806  
; LENGTH: 175  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-28806

Query Match 100.0%; Score 30; DB 4; Length 175;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 PPRGR 5  
Db 127 PPRGR 131

## RESULT 15

US-09-252-991A-29848  
; Sequence 29848, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 29848  
; LENGTH: 177  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-29848

Query Match 100.0%; Score 30; DB 4; Length 177;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 PPRGR 5  
Db 48 PPRGR 52

Search completed: April 6, 2004, 16:19:42  
Job time : 8.33645 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:52:34 ; Search time 19.0654 Seconds  
(without alignments)  
85.771 Million cell updates/sec

Title: US-10-009-709-13

Perfect score: 104

Sequence: 1 GGHPRPRGRPGQPQQ 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR 78.\*

2: PIR1.\*

3: PIR2.\*

4: PIR3.\*

5: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	166	1	PIHUSC
2	104	100.0	166	2	B25372
3	104	100.0	171	2	A27307
4	68	65.4	392	1	PIHUB6
5	66.5	63.9	295	2	B48013
6	65	62.5	76	2	C38355
7	64	61.5	188	2	JH0481
8	62	59.6	117	2	D40750
9	62	59.6	128	2	D38355
10	62	59.6	212	2	B36298
11	62	59.6	251	1	PIHUPF
12	62	59.6	309	2	S10889
13	62	59.6	310	1	PIHUSD
14	60	57.7	206	1	PIRT3
15	59.5	57.2	183	2	A26548
16	58	55.8	170	2	A48013
17	58	55.8	231	2	T27396
18	57.5	55.3	452	1	S14332
19	57	54.8	115	2	A23925
20	57	54.8	438	2	B32654
21	56.5	54.3	797	2	S33590
22	56	53.8	506	2	B56201
23	56	53.8	514	2	A56201
24	55.5	53.4	188	2	D29149
25	55.5	53.4	317	2	A28996
26	55	52.9	485	2	T30190
27	54.5	52.4	300	2	S19560
28	54.5	52.4	301	2	E29149
29	54	51.9	125	2	S40153

30 54 51.9 429 2 JC4965  
31 54 51.9 446 2 T25427  
32 53 51.0 469 2 I37451  
33 53 51.0 707 2 A46302  
34 53 51.0 875 2 T10340  
35 52.5 50.5 55 2 S23770  
36 52.5 50.5 227 2 C29149  
37 52.5 50.5 240 2 B24264  
38 52.5 50.5 240 2 A24264  
39 52.5 50.5 260 2 S22373  
40 52 50.0 381 2 S48049  
41 52 50.0 452 2 A46195  
42 52 50.0 543 1 JC4070  
43 52 50.0 1384 2 T26656  
44 51.5 49.5 1428 2 T08852  
45 51.5 49.5 2715 2 T13049

#### ALIGNMENTS

##### RESULT 1

###### PIHUSC

salivary proline-rich phosphoprotein precursor PRH2 [validated] - human

N/Alternate names: salivary acidic proline-rich protein PRH2

N/Contains: peptide P-C (basic proline-rich peptide IB-8b); proline-rich phosphoprotein

C/Species: Homo sapiens (man)

C/Date: 31-Mar-1981 #sequence\_revision 12-Apr-1996 #text-change 08-Dec-2000

C/Accession: A25372; A19803; B57868; A92277; A94254; A94425; A91954; S02564; S02563; J

R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.

J. Biol. Chem. 260, 11123-11130, 1985

A/Title: Differential RNA splicing and post-translational cleavages in the human saliv

A/Reference number: A92492; MUID:85289325; PMID:2993301

A/Accession: A25372

A/Molecule type: mRNA

A/Residues: 1-166 <MA>

A/Cross-references: GB:K03202; NID:g190481; PIDN:AAA60183.1; PID:g190482

R/Schlesinger, D.H.; Hay, D.I.

Int. J. Pept. Protein Res. 17, 34-41, 1981

A/Title: Primary structure of the active tryptic fragments of human and monkey salivar

A/Reference number: A91757; MUID:81191179; PMID:7228490

A/Accession: A19803

A/Molecule type: protein

A/Residues: 17-46 <SCH>

R/Kim, H.S.; Maeda, N.

J. Biol. Chem. 261, 6712-6718, 1986

A/Title: Structures of two HaeIII-type genes in the human salivary proline-rich protei

A/Reference number: A57868; MUID:86196106; PMID:3009472

A/Accession: B57868

A/Molecule type: DNA

A/Residues: 1-166 <KIN>

A/Cross-references: GB:M13058; NID:g190513; PIDN:AAA98808.1; PID:g190514

R/Wong, R.S.C.; Bennick, A.

J. Biol. Chem. 255, 5943-5948, 1980

A/Title: The primary structure of a salivary calcium-binding proline-rich phosphoprote

A/Reference number: A92277; MUID:80204368; PMID:7380845

A/Accession: A92277

A/Molecule type: protein

A/Residues: 17-19, 'N', 21-166 <WON>

A/Note: the amino-terminal 46 residues are involved with inhibiting hydroxyapatite for

R/Wong, R.S.C.; Hofmann, T.; Bennick, A.

J. Biol. Chem. 254, 4800-4808, 1979

A/Title: The complete primary structure of a proline-rich phosphoprotein from human sa

A/Reference number: A92254; MUID:79173237; PMID:438215

A/Accession: A92254

A/Molecule type: protein

A/Residues: 17-19, 'N', 21-122 <WO2>

R/Schlesinger, D.H.; Hay, D.I.

In Peptides: Structure and Biological Function (Proc. Sixth Amer. Peptide Symp.), Gros

A/Title: Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent

A/Reference number: A94425

RESULT 3  
A27307  
proline-rich phosphoprotein (gene PRH1, Db allele) - human  
NAlternate names: salivary acidic proline-rich protein  
C/Species: Homo sapiens (man)

proline-rich phosphoprotein (gene PRH1, Db allele) - human  
A27307  
N; Alternate names: salivary acidic proline-rich protein  
C; Species: Homo sapiens (man)

C>Date: 30-Jun-1998 #sequence\_revision 30-Jun-1998 #text\_change 29-Aug-1997  
 C/Accession: A27307  
 R/Azen, E.A.; Kim, H.S.; Goodman, P.; Flynn, S.; Maeda, N.  
 Am. J. Hum. Genet. 41, 1035-1047, 1987  
 A>Title: Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich protein  
 A/Reference number: A27307; MUID:188074309; PMID:3697941  
 A/Accession: A27307  
 A/Status: nucleic acid sequence not shown  
 A/Molecule type: DNA  
 A/Residues: 1-171 <AZE>  
 A/Cross-references: EMBL:X03203  
 C/Genetics:  
 A/Gene: GDB:PRH1  
 A/Cross-references: GDB:119515; OMIM:168730  
 A/Map position: 12p13.2-12p13.2  
 C/Superfamily: proline-rich protein  
 C/Keywords: phosphoprotein

Query Match 100.0%; Score 104; DB 2; Length 171;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-05;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRRGRGPGPPQQ 17  
 |||||  
 Db 120 GGHPRRGRGPGPPQQ 136

RESULT 4  
 PIHUB6  
 salivary proline-rich phosphoprotein precursor PRB1 (large allele) [validated] - human  
 N/Contains: peptide IB-1; peptide P-E (peptide IB-9); peptide P-F; peptide P-H  
 C/Species: Homo sapiens (man)  
 C/Dates: 04-Dec-1986 #sequence\_revision 12-Apr-1996 #text\_change 08-Dec-2000  
 C/Accession: B40750; A40750; C25372; S02128; A03293; A90502; A91974; A05  
 R/Azen, E.A.; Latreille, P.; Niece, R.L.  
 Am. J. Hum. Genet. 53, 264-278, 1993  
 A>Title: PRB1 gene variants coding for length and null polymorphisms among human salivary  
 A/Reference number: A40750; MUID:93304421; PMID:8317492  
 A/Accession: B40750  
 A/Molecule type: DNA  
 A/Residues: 35-392 <AZE>  
 A/Cross-references: GB:S62941  
 A/Experimental source: subject C.J. (large allele)  
 A/Accession: C40750  
 A/Molecule type: DNA  
 A/Residues: 35-127, 'R', 129-148, 'R', 150-151, 153-187, 'K', 189-272, 'S', 274-336, 'S', 338-392 <  
 A/Cross-references: GB:S62929  
 A/Experimental source: subject M.V.O. (large allele)  
 A/Accession: A40750  
 A/Molecule type: DNA  
 A/Residues: 35-183, 245-270, 'Q', 272-392 <AZ3>  
 A/Cross-references: GB:S62928  
 A/Experimental source: subject C.J. (medium allele)  
 A/Note: Authors translated the codon CAA for residue 272 as Arg  
 R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
 J. Biol. Chem. 260, 11123-11130, 1985  
 A>Title: Differential RNA splicing and post-translational cleavages in the human salivary  
 A/Reference number: A92492; MUID:85289325; PMID:2993301  
 A/Accession: C25372  
 A/Molecule type: mRNA  
 A/Residues: 1-183, 245-392 <MAE>  
 A/Cross-references: GB:K03204; NID:G190485; PIDN:AAA60185.1; PID:G190486  
 A/Note: Alternatively splice forms lacking portions of the repeat region were also found  
 R/Lyons, K.M.; Stein, J.H.; Smithies, O.  
 Genetics 120, 267-278, 1988  
 A>Title: Length polymorphisms in human proline-rich protein genes generated by intragenic  
 A/Reference number: S02127; MUID:89121440; PMID:2851479  
 A/Accession: S02128  
 A/Status: translation not shown  
 A/Molecule type: DNA  
 A/Residues: 35-127, 250-273, 'R', 275-277, 'R', 279-336, 'S', 338-392 <LYO>  
 A/Cross-references: EMBL:X07517  
 A/Accession: S02127

A/Status: translation not shown  
 A/Molecule type: DNA  
 A/Residues: 35-183, 245-392 <LY2>  
 A/Cross-references: EMBL:X07516  
 R/Kaufman, D.; Hofmann, T.; Bennick, A.; Keller, P.  
 Biochemistry 25, 2387-2392, 1986  
 A>Title: Basic proline-rich proteins from human parotid saliva: complete covalent stru  
 A/Reference number: A90502; MUID:86243355; PMID:3521730  
 A/Accession: A03293  
 A/Molecule type: protein  
 A/Residues: 17-38, 'AP', 41-51, 92-148, 'R', 150-152 <KA2>  
 A/Note: among nine basic proline-rich peptides isolated from the saliva, this peptide  
 A/Accession: A90502  
 A/Molecule type: protein  
 A/Residues: 275-336, 'S', 338-392 <KAU>  
 R/Saitoh, E.; Isemura, S.; Sanada, K.  
 J. Biochem. 94, 1991-1999, 1983  
 A>Title: Further fractionation of basic proline-rich peptides from human parotid saliv.  
 A/Reference number: A91974; MUID:84161824; PMID:6671974  
 A/Contents: P-H  
 A/Accession: A91974  
 A/Molecule type: protein  
 A/Residues: 'S', 338-392 <SAI>  
 R/Azen, E.; Lyons, K.M.; McGonigal, T.; Barrett, N.L.; Clements, L.S.; Maeda, N.; Vani  
 Proc. Natl. Acad. Sci. U.S.A. 81, 5561-5565, 1984  
 A/Reference number: A94005; MUID:84298176; PMID:6089212  
 A/Accession: A05261  
 A/Molecule type: DNA  
 A/Residues: 35-39, 'P', 41-84, 'G', 86, 'R', 87-154, 'R', 218-246; 300-306, 'T', 308-329, 'C', 331-  
 A/Accession: A05262  
 A/Molecule type: DNA  
 A/Residues: 'N', 57-59, 'A', 61-69; 334-336, 'S', 338-339, 'R', 341-392 <AZ5>  
 R/Kaufman, D.; Wong, R.; Bennick, A.; Keller, P.  
 Biochemistry 21, 6558-6562, 1982  
 A>Title: Basic proline-rich proteins from human parotid saliva: complete covalent stru  
 A/Reference number: A90464; MUID:83101329; PMID:6924859  
 A/Contents: IB-9  
 A/Accession: A90464  
 A/Molecule type: protein  
 A/Residues: 92-127, 'R', 129-148, 'R', 150-152 <KA3>  
 R/Isemura, S.; Saitoh, E.; Sanada, K.  
 J. Biochem. 91, 2067-2075, 1982  
 A>Title: Fractionation and characterization of basic proline-rich peptides of human pa.  
 A/Reference number: A91966; MUID:83007119; PMID:7118863  
 A/Contents: P-E  
 A/Accession: A91966  
 A/Molecule type: protein  
 A/Residues: 92-127, 'R', 129-148, 'R', 150-152 <ISE>  
 C/Comment: This peptide contains 21-residue repeats, two of which have internal 7-resid  
 C/Genetics:  
 A/Gene: GDB:PRB1  
 A/Cross-references: GDB:119511; OMIM:180989  
 A/Map position: 12p13.2-12p13.2  
 A/Note: each of the tandem repeats contains a candidate splice acceptor site, and seve  
 C/Superfamily: proline-rich protein  
 C/Keywords: alternative splicing; duplication; parotid gland; phosphoprotein; pyroglut.  
 F17-51/Domain: signal sequence #status predicted <SIG>  
 F17-51/2-152/Product: basic proline-rich peptide IB-1 #status experimental <IB1>  
 F192-152/Product: basic proline-rich peptide P-E #status experimental <PPE>  
 F192-392/Product: basic proline-rich peptide IB-6 #status experimental <PPE6>  
 F192-335/Product: basic proline-rich peptide P-F #status experimental <PPF>  
 F192-392/Product: basic proline-rich peptide P-H #status experimental <PPH>  
 F17/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status experim  
 F124/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 65.4%; Score 68; DB 1; Length 392;  
 Best Local Similarity 61.9%; Pred. No. 0.43;  
 Matches 13; Conservative 2; Mismatches 4; Gaps 1;  
 QY 1 GGHPRRGRGPGPPQQ 17  
 |||||  
 Db 226 GGHPRRGRGPGPPQQ 246

C;Genetics:  
A;introns: 22/1; 34/1; 187/2  
C;Superfamily: proline-rich protein

Query Match 61.5%; Score 64; DB 2; Length 188;  
Best Local Similarity 63.2%; Pred.No. 0.65;  
Matches 12; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 1 GQHPR--PPRGRRPGQPQQ 17  
| : | : | : | : | : |  
Db 105 GNKPGGPPPPGKGPGQPQQ 123

RESULT 8  
D40750  
proline-rich protein PRB1/2S (EA) - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 19-May-1994 #sequence\_revision 19-May-1994 #text\_change 03-May-1996  
C;Accession: D40750  
R;Azzen, E.A.; Latreille, P.; Niece, R.L.  
Am. J. Hum. Genet. 53, 264-278, 1993  
A;Title: PRB1 gene variants coding for length and null polymorphisms among human saliva  
A;Reference number: A40750; MUID:93304421; PMID:8317492  
A;Accession: D40750  
A>Status: Preliminary  
A:Molecule type: DNA  
A;Residues: 1-117 <AZZ>  
A;Cross-references: GB:S62930  
C;Superfamily: proline-rich protein

Query Match 59.6%; Score 62; DB 2; Length 117;  
Best Local Similarity 57.1%; Pred.No. 0.73;  
Matches 12; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

QY 1 GGH----PPRAGRPGQPQQ 17  
| : | : | : | : | : |  
Db 9 GGKPGGPPPPGKGPGGPQQ 29

RESULT 9  
D38355  
basic proline-rich peptide IB-8a - human (fragments)  
C;Species: Homo sapiens (man)  
C;Date: 23-Aug-1991 #sequence\_revision 23-Aug-1991 #text\_change 12-Apr-1995  
C;Accession: D38355  
R;Kauffman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
Biochemistry 30, 3351-3356, 1991  
A;Title: Basic proline-rich proteins from human parotid saliva: relationships of the C  
A;Reference number: A38355; MUID:91190884; PMID:1849422  
A;Accession: D38355  
A>Status: Preliminary  
A:Molecule type: Protein  
A;Residues: 1-128 <NAU>  
C;Superfamily: proline-rich protein

Query Match 59.6%; Score 62; DB 2; Length 128;  
Best Local Similarity 57.1%; Pred.No. 0.79;  
Matches 12; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

QY 1 GGH----PPRGRGRRPGQPQQ 17  
| : | : | : | : | : |  
Db 13 GGNQGCGPPPPGKGPGGPQQ 33

RESULT 10  
B36298  
proline-rich protein PRB3S (cys) - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 28-Mar-1991 #sequence\_revision 28-Mar-1991 #text\_change 29-Aug-1997  
C;Accession: B36298  
R;Azzen, E.A.; Minaguchi, K.; Latreille, P.; Kim, H.S.  
Am. J. Hum. Genet. 47, 686-697, 1990  
A;Title: Alleles at the PRB3 locus coding for a disulfide-bonded human salivary prolin



A/Reference number: A36298; MUID:91022705; PMID:2171329  
 A/Accession: B36298  
 A/Status: preliminary; not compared with conceptual translation  
 A/Molecule type: DNA  
 A/Residues: 1-212 <AZE>  
 C/Genetics:  
 A/Gene: GDB:PRB3  
 A/Cross-references: GDB:119513; OMIM:168840  
 A/Map position: 12p13.2-12p13.2  
 C/Superfamily: proline-rich protein

Query Match 59.6%; Score 62; DB 2; Length 212;  
 Best Local Similarity 57.9%; Pred. No. 1.2;  
 Matches 11; Conservative 3; Mismatches 1; Indels 4; Gaps 1;

QY 1 GGHRP-----PPRRPQGGPP 15  
 |||: |||: |||: |||:  
 Db 176 GGNPQQPFPFPAGKPGQPP 194

## RESULT 11

PIHUPF  
 salivary proline-rich glycoprotein precursor PRB2 [validated] - human (fragment)  
 N/Alternate names: basic proline-rich peptide IB-8c precursor; proline-rich protein (cld  
 N/Contains: basic proline-rich peptide IB-4; basic proline-rich peptide P-F  
 C/Species: Homo sapiens (man)  
 C/Date: 15-Nov-1984 #sequence revision 12-Apr-1996 #text change 08-Dec-2000  
 C/Accession: E25372; A60827; A38355; A38355; F38355  
 R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
 J. Biol. Chem. 260, 1123-1130, 1985  
 A/Title: Differential RNA splicing and post-translational cleavages in the human salivary  
 A/Reference number: A92492; MUID:85289325; PMID:2993301  
 A/Accession: E25372  
 A/Molecule type: mRNA  
 A/Residues: 1-251 <MAE>  
 A/Cross-references: GB:K03208; NID:gl90509; PIDN:AAA60189.1; PID:gl90510  
 R/Mamula, P.W.; Morley, D.J.; Larsen, S.H.; Karn, R.C.  
 Biochem. Genet. 26, 165-175, 1988  
 A/Title: Expression of human salivary protein genes.  
 A/Reference number: A60827; MUID:88240287; PMID:3288192  
 A/Accession: A60827  
 A/Status: not compared with conceptual translation  
 A/Molecule type: mRNA  
 A/Residues: 175-251 <MAE>  
 R/Saitoh, E.; Isemura, S.; Sanada, K.  
 J. Biochem. 93, 883-888, 1983

A/Title: Complete amino acid sequence of a basic proline-rich peptide, P-F, from human P  
 A/Reference number: A03294; MUID:83265674; PMID:6874669  
 A/Accession: A03294

A/Molecule type: protein  
 A/Residues: 134-194 <SAI>  
 A/Experimental source: saliva  
 R/Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
 Biochemistry 30, 3351-3356, 1991  
 A/Title: Basic proline-rich proteins from human parotid saliva: relationships of the cov  
 A/Reference number: A38355; MUID:91190884; PMID:1849422  
 A/Accession: B38355  
 A/Molecule type: protein  
 A/Residues: 134-194 <KAU>  
 A/Experimental source: saliva  
 A/Note: this peptide, which is closely related to that of peptide P-E, contains three 21  
 A/Accession: A38355  
 A/Molecule type: protein  
 A/Residues: 10-67, 'R' <KA2>  
 A/Accession: F38355  
 A/Molecule type: protein  
 A/Residues: 196-251 <KA3>  
 C/Genetics:  
 A/Gene: GDB:PRB2  
 A/Cross-references: GDB:119512; OMIM:168810  
 A/Map position: 12p13.2-12p13.2  
 C/Superfamily: proline-rich protein  
 C/Keywords: glycoprotein; saliva; tandem repeat

F:134-194/Product: basic proline-rich peptide P-F #status experimental <MAT1>  
 F:196-251/Product: basic proline-rich peptide IB-4 #status experimental <MAT2>  
 F:3,65,107/Binding site: carbohydrate (Aen) (covalent) #status predicted

Query Match 59.6%; Score 62; DB 1; Length 251;  
 Best Local Similarity 55.0%; Pred. No. 1.5;  
 Matches 11; Conservative 4; Mismatches 1; Indels 4; Gaps 1;

QY 1 GGHRP-----PPRRPQGGPPQ 16  
 |||: |||: |||: |||:  
 Db 220 GGNPQQPFPFPAGQPGQPPR 239

## RESULT 12

SI0889  
 proline-rich protein - human  
 C/Species: Homo sapiens (man)  
 C/Date: 07-Oct-1994 #sequence\_revision 26-May-1995 #text\_change 20-Aug-1999  
 C/Accession: SI0889  
 R/Lyons, K.M.; Stein, J.H.; Smithies, O.  
 Genetics 120, 267-278, 1988  
 A/Title: Length polymorphisms in human proline-rich protein genes generated by intrage  
 A/Reference number: S02127; MUID:89121440; PMID:2851479  
 A/Accession: SI0889  
 A/Status: preliminary; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-309 <LYO>  
 A/Cross-references: EMBL:X07881; NID:g35637; PIDN:CAA30728.1; PID:g295669  
 C/Genetics:  
 A/Introns: 22/1, 34/1  
 C/Superfamily: proline-rich protein

Query Match 59.6%; Score 62; DB 2; Length 309;  
 Best Local Similarity 57.9%; Pred. No. 1.8;  
 Matches 11; Conservative 3; Mismatches 1; Indels 4; Gaps 1;

QY 1 GGHRP-----PPRRPQGGPP 15  
 |||: |||: |||: |||:  
 Db 273 GGNPQQPFPFPAGKPGQPP 291

## RESULT 13

PIHUSD  
 salivary proline-rich glycoprotein precursor PRB4 (large allele) [validated] - human  
 N/Contains: basic proline-rich protein IB-5; proline-rich peptide P-D  
 C/Species: Homo sapiens (man)  
 C/Date: 19-Feb-1984 #sequence\_revision 12-Apr-1996 #text\_change 08-Dec-2000  
 C/Accession: S03176; S03175; S10890; D25372; S38355; A03295; A61294; S62891  
 R/Lyons, K.M.; Stein, J.H.; Smithies, O.  
 Genetics 120, 267-278, 1988  
 A/Title: Length polymorphisms in human proline-rich protein genes generated by intrage  
 A/Reference number: S02127; MUID:89121440; PMID:2851479  
 A/Accession: S03176  
 A/Status: translation not shown  
 A/Molecule type: DNA  
 A/Residues: 35-310 <LY1>  
 A/Cross-references: EMBL:X07715  
 A/Note: large allele  
 A/Accession: S03175  
 A/Status: translation not shown  
 A/Molecule type: DNA  
 A/Residues: 35-36, 'E', 38-112, 155-310 <LY2>  
 A/Cross-references: EMBL:X07704  
 A/Note: medium allele  
 A/Accession: SI0890  
 A/Status: preliminary; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-38, 60-112, 'T', 114-115, 'P', 117-121, 185-271, 'A', 273-310 <LY3>  
 A/Cross-references: EMBL:X07882; NID:g35647; PIDN:CAA30729.1; PID:g295670  
 R/Maeda, N.; Kim, H.S.; Azen, E.A.; Smithies, O.  
 J. Biol. Chem. 260, 1123-1130, 1985  
 A/Title: Differential RNA splicing and post-translational cleavages in the human saliv  
 A/Reference number: A92492; MUID:85289325; PMID:2993301

A:Accession: D25372  
A:Molecule type: mRNA  
A:Residues: 1-36, 'E', 38-112, 'T', 114-115, 'P', 117-121, 185-271, 'A', 273-310 <MAE>  
R:Kaufman, D.L.; Bennick, A.; Blum, M.; Keller, P.J.  
Biochemistry 30, 3351-3356, 1991  
A:Title: Basic proline-rich proteins from human parotid saliva: relationships of the cov  
A:Reference number: A38355; MUID:91190884; PMID:1849422  
A:Accession: E38355  
A:Molecule type: protein  
A:Residues: 241-254, 'KN', 257-310 <XAU>  
R:Saitoh, E.; Isemura, S.; Sanada, K.  
J. Biochem. 93, 495-502, 1983  
A:Title: Complete amino acid sequence of a basic proline-rich peptide, P-D, from human P  
A:Reference number: A03295; MUID:83186122; PMID:6841349  
A:Accession: A03295  
A:Molecule type: protein  
A:Residues: 241-310 <SAI>  
R:Shimomura, H.; Kanai, Y.; Sanada, K.  
J. Biochem. 93, 857-863, 1983  
A:Title: Amino acid sequences of glycopeptides obtained from basic proline-rich glycopro  
A:Reference number: A61294; MUID:83265671; PMID:6874667  
A:Accession: A61294  
A:Molecule type: protein  
A:Residues: 54-57, 'E', 59-73, 'R', 82-101 <SHI>  
R:Charlton, A.J.; Baxter, N.J.; Lilley, T.H.; Haslam, E.; McDonald, C.J.; Williamson, M.  
FEBS Lett. 382, 289-292, 1996  
A:Title: Tannin interactions with a full-length human salivary proline-rich protein dis  
A:Reference number: S62891; MUID:96184506; PMID:8605987  
A:Accession: S62891  
A:Molecule type: protein  
A:Residues: 241-252 <CHA>  
A:Note: amino end of peptide designated basic proline-rich protein IB-5  
A:Note: it is unclear from the peptide sequence whether this is a product of the PRB2 (R  
C:Genetics:  
A:Gene: GDB:PRB4  
A:Cross-references: GDB:119514; OMIM:180990  
A:Map position: 12p13.2-12p13.2  
A:Introns: 22/1; 34/1  
A:Note: the list of introns may be incomplete  
C:Superfamily: proline-rich protein  
C:Keywords: glycoprotein; saliva; tandem repeat  
F:1-16/Domain: signal sequence #status predicted <SIG>  
F:241-310/Product: proline-rich peptide P-D #status experimental <MAT>  
F:66,87,171/Binding site: carbohydrate (Asn) (covalent) #status experimental  
F:108,150,192,213,234/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 59.6%; Score 62; DB 1; Length 310;  
Best Local Similarity 57.9%; Pred. No. 1.8;  
Matches 11; Conservative 3; Mismatches 1; Indels 4; Gaps 1;  
QY 1 GGHPR----PPRGPPGPP 15  
DB 273 GGNPQQPQAPPAGKPGQPP 291

RESULT 14  
P1RT3  
acidic proline-rich protein precursor - rat  
N/Alternate names: PRP  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 22-Jun-1999  
C:Accession: A03296  
R:Ziemer, M.A.; Swain, W.F.; Rutter, W.J.; Clements, S.; Ann, D.K.; Carlson, D.M.  
J. Biol. Chem. 259, 10475-10480, 1984  
A:Title: Nucleotide sequence analysis of a proline-rich protein cDNA and peptide homolog  
A:Reference number: A03296; MUID:84289443; PMID:6547951  
A:Accession: A03296  
A:Molecule type: mRNA  
A:Residues: 1-206 <ZIE>  
A:Cross-references: GB:K02247; NID:G206395; PID:AAA41949.1; PID:G206396  
C:Comment: This protein contains six 18- to 19-residue repeats.  
C:Comment: This protein may protect teeth by binding to tannins.  
C:Superfamily: proline-rich protein

C:Keywords: duplication; parotid gland; saliva; tandem repeat  
F:1-13/Domain: signal sequence #status predicted <SIG>  
F:14-206/Product: acidic proline-rich protein #status predicted <MAT>  
F:180-189/Region: 18-residue repeats

Query Match 57.7%; Score 60; DB 1; Length 206;  
Best Local Similarity 54.7%; Pred. No. 2.1;  
Matches 11; Conservative 3; Mismatches 1; Indels 2; Gaps 1;  
QY 2 GGHPR--PPRGPPGPPQ 16  
DB 111 GNPQPPPPGPPGPPQ 127

RESULT 15  
A26548  
acidic proline-rich protein H29 precursor - Golden hamster  
C:Species: Mesocricetus auratus (golden hamster)  
C>Date: 19-Nov-1988 #sequence\_revision 19-Nov-1988 #text\_change 15-Nov-1996  
C:Accession: A26548  
R:Ann, D.K.; Gadbois, D.; Carlson, D.M.  
J. Biol. Chem. 262, 3958-3963, 1987  
A:Title: Structure, organization, and regulation of a hamster proline-rich protein gen  
A:Reference number: A26548; MUID:87165996; PMID:3031057  
A:Accession: A26548  
A:Molecule type: DNA  
A:Residues: 1-183 <ANN>  
A:Experimental source: parotid gland  
C:Superfamily: proline-rich protein  
F:1-14/Domain: signal sequence #status predicted <SIG>  
F:15-183/Product: acidic proline-rich protein H29 #status predicted <MAT>

Query Match 57.2%; Score 59.5; DB 2; Length 183;  
Best Local Similarity 65.0%; Pred. No. 2.2;  
Matches 13; Conservative 1; Mismatches 3; Indels 3; Gaps 2;  
QY 1 GGH--PRPPR-GRPGPPQ 17  
DB 75 GDHQRPRPFKPNQGGPPQ 94

Search completed: April 6, 2004, 16:16:56  
Job time: 20.0654 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:31:19 ; Search time 10.9626 Seconds  
(without alignments)  
80.746 Million cell updates/sec

Title: US-10-009-709-13  
Perfect score: 104  
Sequence: 1 GGHPRPPRGPPQPPQ 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	166	1 PRP1_HUMAN	P02810 homo sapien
2	68	65.4	331	1 PRP2_HUMAN	P04280 homo sapien
3	62	59.6	61	1 PRP3_HUMAN	P02811 homo sapien
4	62	59.6	96	1 PRP4_HUMAN	P04281 homo sapien
5	62	59.6	174	1 PRP5_HUMAN	P81489 homo sapien
6	62	59.6	234	1 PRP6_HUMAN	P10161 homo sapien
7	62	59.6	247	1 PRP7_HUMAN	P10163 homo sapien
8	62	59.6	251	1 PRP8_HUMAN	P02812 homo sapien
9	62	59.6	276	1 PRP9_HUMAN	P10162 homo sapien
10	60	57.7	206	1 PRP10_HUMAN	P04474 rattus norv
11	59.5	57.2	183	1 PRP11_HUMAN	P06680 mesocricetu
12	57.5	55.3	439	1 PRP12_HUMAN	Q17094 loligo subu
13	57.5	55.3	452	1 PRP13_HUMAN	P24603 loligo forb
14	57	54.8	438	1 PRP14_HUMAN	Q9yeb2 aeropyrum p
15	56.5	54.3	797	1 PRP15_HUMAN	P25644 saccharomyc
16	56	53.8	514	1 PRP16_HUMAN	Q63943 mus musculu
17	54	51.9	125	1 PRP17_HUMAN	P47194 loligo peal
18	54	51.9	429	1 PRP18_HUMAN	P41969 mus musculu
19	53	51.0	469	1 PRP19_HUMAN	P55316 homo sapien
20	53	51.0	656	1 PRP20_HUMAN	Q60865 mus musculu
21	53	51.0	707	1 PRP21_HUMAN	P23246 homo sapien
22	53	51.0	875	1 PRP22_HUMAN	Q83949 orgyia pseu
23	53	51.0	1860	1 PRP23_HUMAN	Q90940 homo sapien
24	53	51.0	1983	1 PRP24_HUMAN	Q90948 mus musculu
25	52.5	50.5	261	1 PRP25_HUMAN	P05142 mus musculu
26	52.5	50.5	296	1 PRP26_HUMAN	P05143 mus musculu
27	52	50.0	452	1 PRP27_HUMAN	P30553 rattus norv
28	52	50.0	543	1 PRP28_HUMAN	P54739 streptomyce
29	51.5	49.5	2716	1 PRP29_HUMAN	Q8in94 drosophila
30	51.5	49.5	2911	1 PRP30_HUMAN	P35556 homo sapien
31	51	49.0	161	1 PRP31_HUMAN	P04487 herpes simp
32	51	49.0	161	1 PRP32_HUMAN	P06958 herpes simp
33	51	49.0	333	1 PRP33_HUMAN	P19198 dictyosteli

RESULT 1	PRPC_HUMAN	STANDARD;	PRT;	166 AA.
AC	P02810;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Salivary acidic proline-rich phosphoprotein 1/2 precursor (PRP-1/PRP-3) (PRP-2/PRP-4) (PIF-F/PIF-S) (Protein A/protein C) [Contains: Peptide P-C].			
DE	Peptide P-C.			
GN	PRH1 AND PRH2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ALLELES PRH1-4 AND PRH2-1).			
RX	MEDLINE=86196106; PubMed=3009472;			
RA	Kim H.-S., Maeda N.;			
RT	"Structures of two HaerII-type genes in the human salivary proline-rich protein multigene family.";			
RL	J. Biol. Chem. 261:6712-6718(1986).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ALLELES PRH2-1 AND PRH1-PIF).			
RX	MEDLINE=85289325; PubMed=2393301;			
RA	Maeda N., Kim H.-S., Azen E.A., Smithies O.;			
RT	"Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system.";			
RL	J. Biol. Chem. 260:11123-11130(1985).			
RN	[3]			
RP	SEQUENCE OF 17-166 (PRP-1 TO PRP-4; PIF-F AND PIF-S).			
RX	MEDLINE=89061650; PubMed=3196309;			
RA	Hay D.I., Bennick A., Schlesinger D.H., Minaguchi K.,			
RT	"The primary structures of six human salivary acidic proline-rich proteins (PRP-1, PRP-2, PRP-3, PRP-4, PIF-S and PIF-F).";			
RL	Biochem. J. 255:15-21(1988).			
RN	[4]			
RP	SEQUENCE OF 17-166 FROM N.A. (PA ALLELE).			
RX	MEDLINE=88074309; PubMed=3687941;			
RA	Azen E.A., Kim H.-S., Goodman P., Flynn S., Maeda N.;			
RT	"Alleles at the PRH1 locus coding for the human salivary-acidic proline-rich proteins Pa, Db, and PIF.";			
RL	Am. J. Hum. Genet. 41:1035-1047(1987).			
RN	[5]			
RP	SEQUENCE OF 17-166 (PRP-2).			
RX	MEDLINE=86222916; PubMed=3710693;			
RA	Schlesinger D.H., Hay D.I.;			
RT	"Complete covalent structure of a proline-rich phosphoprotein, PRP-2, an inhibitor of calcium phosphate crystal growth from human parotid saliva.";			
RL	Int. J. Pept. Protein Res. 27:373-379(1986).			
RN	[6]			
RP	SEQUENCE OF 17-166 (PROTEIN C).			
RX	MEDLINE=80204368; PubMed=7380845;			
RA	Wong R.S.C., Bennick A.;			

P70460 mus musculu  
P50552 homo sapien  
P50551 canis faml  
Q14684 homo sapien  
Q92296 mus musculu  
Q9ubv2 homo sapien  
P21519 drosophila  
Q14686 homo sapien  
P32239 homo sapien  
Q92519 mycobacteri  
O60885 homo sapien  
P48634 homo sapien

RT "The primary structure of a salivary calcium-binding proline-rich phosphoprotein (protein C), a possible precursor of a related salivary protein A.";  
 RL J. Biol. Chem. 255:5943-5948(1980).  
 [7]  
 RN SEQUENCE OF 17-46 (PROTEIN C).  
 RX MEDLINE=81191179; PubMed=7228490;  
 RA Schlesinger D.H., Hay D.I.;  
 RT "Primary structure of the active tryptic fragments of human and monkey salivary anionic proline-rich proteins.";  
 RL Int. J. Pept. Protein Res. 17:34-41(1981).  
 [8]  
 RN SEQUENCE OF 17-122 (PROTEIN A).  
 RX MEDLINE=79173237; PubMed=438215;  
 RA Wong R.S.C., Hofmann T., Bennis A.;  
 RT "The complete primary structure of a proline-rich phosphoprotein from human saliva.";  
 RL J. Biol. Chem. 254:4800-4808(1979).  
 [9]  
 RN SEQUENCE OF 17-122 (PROTEIN A).  
 RX Schlesinger D.H., Hay D.I.;  
 RT "Complete primary structure of a proline-rich phosphoprotein (PRP-4), a potent inhibitor of calcium phosphate precipitation in human parotid saliva.";  
 RL (In) Gross E., Meienhofer J. (eds.);  
 RL Peptides: structure and biological function (Proceedings of the 6th American peptide symposium), pp.133-136, Pierce Chemical Co., Rockford IL. (1979).  
 [10]  
 RN SEQUENCE OF 123-166 (PEPTIDE P-C).  
 RX MEDLINE=80227634; PubMed=7390979;  
 RA Isemura S., Saitoh E., Sanada K.;  
 RT "The amino acid sequence of a salivary proline-rich peptide, P-C, and its relation to a salivary proline-rich phosphoprotein, protein C.";  
 RL J. Biochem. 87:1071-1077(1980).  
 [11]  
 RN VARIANT PRH2-3 LYS-163.  
 RA Azen E.A.;  
 RT "A frequent mutation in the acidic proline-rich protein gene, PRH2, causing a Q147K change closely adjacent to the bacterial binding domain of the cognate salivary PRP (Pr1) in Afro-Americans.";  
 RL Hum. Mutat. 12:72-72(1998).  
 CC -1- FUNCTION: PRP's act as highly potent inhibitors of crystal growth of calcium phosphates. They provide a protective and reparative environment for dental enamel which is important for the integrity of the teeth.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- PM: Proteolytically cleaved: PRP-2, PRP-1, and PIF-S yield PRP-4, PRP-3 (protein A), and PIF-F, respectively.  
 CC -1- POLYMORPHISM: Allele PRH1-4 is also known as PRH1 DA allele; allele PRH2-1 is also known as PR1 or protein C; allele PRH2-3 is also known as PR1'.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: K03202; AAA60183.1;  
 CC EMBL: K03203; AAA60184.1;  
 CC EMBL: M13057; AAA98807.1;  
 CC EMBL: M13058; AAA98808.1;  
 CC Genbank: HGNC:9366; PRH1.  
 CC Genbank: HGNC:9367; PRH2.  
 CC MIM: 168730;  
 CC MIM: 168790;  
 CC MIM: 168710;  
 CC GO: GO:0005615; C:extracellular space; TAS.  
 CC Repeat; Parotid gland; Phosphorylation; Signal; Polymorphism;  
 KW Pyrolidone carboxylic acid.

FT SIGNAL 1 16  
 FT CHAIN 17 166  
 FT CHAIN 17 122  
 FT CHAIN 17 122  
 FT CHAIN 123 166  
 FT DOMAIN 17 46  
 FT MOD\_RES 17 17  
 FT MOD\_RES 24 24  
 FT MOD\_RES 38 38  
 FT VARIANT 20 20  
 FT VARIANT 66 66  
 FT VARIANT 153 163  
 FT CONFLICT 41 41  
 FT SEQUENCE 166 AA; 17017 MW; A7DF62BF94E3C3EF CRC64;  
 SQ  
 Query Match 100.0%; Score 104; DB 1; Length 166;  
 Best Local Similarity 100.0%; Pred. No. 9.7e-06;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGHPRPPRGPPGPPQ 17  
 DB 115 GGHPRPPRGPPGPPQ 131  
 RESULT 2  
 ID PRP1\_HUMAN STANDARD; PRT; 331 AA.  
 AC P04280;  
 DT 20-04-1987 (Rel. 04; Created)  
 DT 13-AUG-1987 (Rel. 05; Last sequence update)  
 DT 28-FEB-2003 (Rel. 41; Last annotation update)  
 DE Salivary proline-rich protein precursor (Clones CP3, CP4 and CP5)  
 DE [Contains: Basic peptide IB-6; Peptide P-H].  
 GN PRB1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=85289325; PubMed=2993301;  
 RA Maeda N., Kim H.-S., Azen E.A., Smithies O.;  
 RT "Differential RNA splicing and post-translational cleavages in the human salivary proline-rich protein gene system.";  
 RL J. Biol. Chem. 260:11123-11130(1985).  
 [2]  
 RP SEQUENCE OF 214-331.  
 RX MEDLINE=86243355; PubMed=3521730;  
 RA Kauffman D., Hofmann T., Bennis A., Keller P.;  
 RT "Basic proline-rich proteins from human parotid saliva: complete covalent structures of proteins IB-1 and IB-6.";  
 RL Biochemistry 25:2387-2392(1986).  
 [3]  
 RP SEQUENCE OF 276-331.  
 RX MEDLINE=84161824; PubMed=6671974;  
 RA Saitoh E., Isemura S., Sanada K.;  
 RT parotid saliva and complete amino acid sequence of basic proline-rich peptide P-H.";  
 RL J. Biochem. 94:1991-1997(1983).  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC

```

CC -----
DR EMBL; K03204; AAA60185.1; -
DR EMBL; K03205; AAA60186.1; -
DR EMBL; K03206; AAA60187.1; -
DR PIR; B40750; PIHUB6.
DR Genew; HGNC:9337; FBEL.
DR MIM; 180989; -
KW Repeat; Parotid gland; Multigene family; Signal.
FT SIGNAL 1 16 PEPTIDE IB-6.
FT CHAIN 214 331 PEPTIDE P-H.
FT CHAIN 276 331 Missing (in clone CP-4).
FT VARIANT 106 238 /FTID=VAR 005561.
FT VARIANT 106 258 Missing (in clone CP-5).
FT VARIANT 276 276 /FTID=VAR 005562.
FT CONFLICT 276 276 A -> S (IN REF. 2 AND 3).
FT SEQUENCE 331 AA; 32596 MW; 3F481F08BA39751 CRC64;

Query Match 65.4%; Score 68; DB 1; Length 331;
Best Local Similarity 61.9%; Pred. No. 0.26;
Matches 13; Conservative 2; Mismatches 2; Indels 4; Gaps 1;

QY 1 GGH----PRPRGRPGQPPQ 17
DB 165 GGNQPGPPPPPKKQGGPPQ 185

RESULT 3
PRPP HUMAN
ID PRPP HUMAN STANDARD; PRT; 61 AA.
AC P02811;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Basic proline-rich peptide P-E (IB-9).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP TISSUE=Saliva;
RC MEDLINE=63101329; PubMed=6924859;
RA Kaufman D., Wong R., Bannick A., Keller P.;
RT "Basic proline-rich proteins from human parotid saliva: complete
covalent structure of protein IB-9 and partial structure of protein
IB-6, members of a polymorphic pair."
RL Biochemistry 21:6558-6562(1982).
RN [2]
RP SEQUENCE.
RC TISSUE=Saliva;
RC MEDLINE=83007119; PubMed=7118863;
RA Isemura S., Saitoh E., Sanada K.;
RT "Fractionation and characterization of basic proline-rich peptides of
human parotid saliva and the amino acid sequence of proline-rich
peptide P-E."
RL J. Biochem. 91:2067-2075(1982).
CC -!- MISCELLANEOUS: Peptides IB-9 and P-E are the same peptide.
DR PIR; B40750; PIHUB6.
DR GO; GO:0005576; C:extracellular; NAS.
KW Repeat; Parotid gland.
FT REPEAT 1 21 APPROXIMATE.
FT REPEAT 22 42 APPROXIMATE.
FT REPEAT 43 61 APPROXIMATE.
FT SEQUENCE 61 AA; 6024 MW; F883F78617E05327 CRC64;

Query Match 59.6%; Score 62; DB 1; Length 61;
Best Local Similarity 57.1%; Pred. No. 0.25;
Matches 12; Conservative 3; Mismatches 3; Indels 4; Gaps 1;

QY 1 GGH----PRPRGRPGQPPQ 17
DB 165 GGNQPGPPPPPKKQGGPPQ 185

RESULT 4
PRPP HUMAN
ID PRPP HUMAN STANDARD; PRT; 96 AA.
AC P04281;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Basic proline-rich peptide IB-1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE.
RC MEDLINE=86243355; PubMed=3521730;
RA Kaufman D., Hofmann T., Bannick A., Keller P.;
RT "Basic proline-rich proteins from human parotid saliva: complete
covalent structures of proteins IB-1 and IB-6."
RL Biochemistry 25:2387-2392(1986).
CC -!- DOMAIN: Contains 21-residue repeats, two of which have internal
7-residue repeats.
CC -!- MISCELLANEOUS: There are nine basic proline-rich peptides isolated
from the saliva; this peptide is designated IB-1.
DR PIR; B40750; PIHUB6.
DR KW Phosphorylation; Repeat; Parotid gland; Pyrrolidone carboxylic acid.
FT MOD_RES 1 1 PYRROLIDONE CARBOXYLIC ACID.
FT MOD_RES 8 8 PHOSPHORYLATION.
FT REPEAT 15 35 APPROXIMATE.
FT REPEAT 36 56 APPROXIMATE.
FT REPEAT 57 77 APPROXIMATE.
FT SEQUENCE 96 AA; 9530 MW; FF3B33F68DE104F8 CRC64;

Query Match 59.6%; Score 62; DB 1; Length 96;
Best Local Similarity 57.1%; Pred. No. 0.39;
Matches 12; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

QY 1 GGH----PRPRGRPGQPPQ 17
DB 48 GGNQPGPPPPPKKQGGPPQ 68

RESULT 5
PRPP HUMAN
ID PRPP HUMAN STANDARD; PRT; 174 AA.
AC P81489;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Salivary proline-rich protein II-1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP TISSUE=Saliva;
RC TISSUE=Saliva;
RC MEDLINE=93385383; PubMed=8373986;
RA Kaufman D.L., Keller P.J., Bannick A., Blum M.;
RT "Alignment of amino acid and DNA sequences of human proline-rich
proteins."
RL Crit. Rev. Oral Biol. Med. 4:287-292(1993).
DR GO; GO:0005576; C:extracellular; NAS.
DR InterPro; IPR000637; AT hook.
DR PRINTS; PR00929; ATHOOK.
KW Repeat; Parotid gland; Multigene family.
FT NON_TER 174 174
FT SEQUENCE 174 AA; 17802 MW; D645F106EB1BBSBE CRC64;

Query Match 59.6%; Score 62; DB 1; Length 174;
Best Local Similarity 57.1%; Pred. No. 0.68;
Matches 12; Conservative 2; Mismatches 3; Indels 4; Gaps 1;

```



-----  
 ENBL; X07715; CAA30543.1; ALT\_SEQ.  
 PIR; S03176; PIHUSD.  
 -----

[illegible]

```
Query Match          57.7%; Score 60; DB 1; Length 206;
Best Local Similarity 64.7%; Pred. No. 1.4;
Matches 11; Conservative 3; Mismatches 2; Gaps 1;

OY 2 GHPR--PPRGPPQPPQ 16
Db 111 GNPQPPPPQGGPQPPQ 127

RESULT 11
PRPH MESAU
ID PRPH MESAU STANDARD; PRT; 183 AA.
AC P06680;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1989 (Rel. 38, Last annotation update)
DE Acidic proline-rich protein HP43A precursor.
GN H29.

OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
CX NCBI_TaxID=10036;
RN [1]_TaxID=10036;
RP SEQUENCE FROM N.A.
RX MEDLINE=87165996; PubMed=3031057;
RA Ann D.K., Gadbois D., Carlson D.M.;
RT "Structure, organization, and regulation of a hamster proline-rich
RT protein gene. A multigene family";
RL J. Biol. Chem. 262:3958-3963(1987).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02686; -; NOT_ANNOTATED_CDS.
DR PIR; A26548; A26548.
DR Repeat; Signal.
KW SIGNAL 1 14
FT CHAIN 15 183 ACIDIC PROLINE-RICH PROTEIN HP43A.
FT REPEAT 81 100 1.
FT REPEAT 101 120 2.
FT REPEAT 121 140 3.
FT REPEAT 141 160 4.
FT REPEAT 161 180 5.
SQ SEQUENCE 183 AA; 19774 MW; DDC1BF196AFICA07 CRC64;

Query Match          57.2%; Score 59.5; DB 1; Length 183;
Best Local Similarity 65.0%; Pred. No. 1.4;
Matches 13; Conservative 1; Mismatches 3; Indels 3; Gaps 2;

OY 1 GGH--PPRR--GRPGPPQ 17
Db 75 GDHQKRPKPGNQGPQ 94

RESULT 12
OPSD LOLSU
ID OPSD LOLSU STANDARD; PRT; 439 AA.
AC Q17094;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Rhodopsin (Fragment).
GN RHO.
OS Loligo subulata (Squid) (Alloteuthis subulata).
OC Eukaryota; Metazoa; Mollusca; Cephalopoda; Coleoidea; Neocoleoidea;
OC Decapodiformes; Loliginidae; Loligo.
```

```
NCBI_TaxID=54069;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Epidermis;
RX MEDLINE=94151381; PubMed=8108455;
RA Morris A., Bowmaker J.K., Hunt D.M.;
RT "The molecular basis of a spectral photic environments.";
RL Proc. R. Soc. Lond. B, Biol. Sci. 254:233-240(1993).
CC -!- FUNCTION: VISUAL PIGMENTS ARE THE LIGHT-ABSORBING MOLECULES THAT
CC MEDIATE VISION. THEY CONSIST OF AN APOPROTEIN, OPSIN, COVALENTLY
CC LINKED TO CIS-RETINAL. THIS RECEPTOR IS COUPLED TO THE ACTIVATION
CC OF PHOSPHOLIPASE C.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- PTM: Some or all of the carboxyl-terminal Ser or Thr residues may
CC be phosphorylated.
CC -!- MISCELLANEOUS: THIS RHODOPSIN HAS AN ABSORPTION MAXIMUM AT 499 NM.
CC -!- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC Opsin subfamily.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z49108; CAA88923.1; -;
DR HSP; P02699; 1F88.
DR InterPro; IPR000276; GPCR_Rhodopsn.
DR InterPro; IPR001760; Opsin.
DR InterPro; IPR006031; XYPX.
DR Pfam; PF00001; 7tm1.1.
DR PRINTS; PR02162; XYPX; 8.
DR PRINTS; PR00237; GPCR_RHODOPSIN.
DR PROSITE; PS00237; G PROTEIN RECEPTOR FL1; 1.
DR PROSITE; PS00262; G PROTEIN RECEPTOR FL2; 1.
DR PROSITE; PS00238; OPSIN; 1.
DR Phosphorylation; Retinal protein; Transmembrane; Glycoprotein; Vision;
KW Phosphorylation; lipoprotein; Palmitate; G-protein coupled receptor.
FT NON TER 1 1
FT DOMAIN <1 26 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 27 51 1 (POTENTIAL).
FT DOMAIN 52 63 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 64 90 2 (POTENTIAL).
FT DOMAIN 91 104 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 105 124 3 (POTENTIAL).
FT DOMAIN 125 144 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 145 168 4 (POTENTIAL).
FT DOMAIN 169 192 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 193 220 5 (POTENTIAL).
FT DOMAIN 221 254 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 255 278 6 (POTENTIAL).
FT DOMAIN 279 287 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 288 311 7 (POTENTIAL).
FT DOMAIN 312 >439 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 1 1 N-LINKED (GLCNAC...) (PROBABLE).
FT DISULFID 101 179 BY SIMILARITY.
FT BINDING 298 298 RETINAL CHROMOPHORE (BY SIMILARITY).
FT LIPID 329 329 S-palmitoyl cysteine (By similarity).
FT LIPID 330 330 S-palmitoyl cysteine (By similarity).
FT DOMAIN 363 373 MET-RICH.
FT DOMAIN 374 436 GLN/PRO-RICH.
FT NON TER 439 439
SQ SEQUENCE 439 AA; 49017 MW; B8167DFD8A00390E CRC64;

Query Match          55.3%; Score 57.5; DB 1; Length 439;
Best Local Similarity 64.7%; Pred. No. 5.4;
Matches 11; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

OY 2 GHPRPPRG--RPGPPQ 17
Db 75 GDHQKRPKPGNQGPQ 94
```



[illegible][illegible]

Search completed: April 6, 2004, 16:08:02  
Job time : 11.9626 secs

Query Match 54.8%; Score 57; DB 1; Length 438;  
Best Local Similarity 62.5%; Pred. No. 6.1;  
Matches 10; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 GGHPRPPRGRCGPPQ 16  
DB 6 GGRPPRGFRDPPE 21

RESULT 15  
PAT1 YEAST  
ID PAT1 YEAST STANDARD; PRT; 797 AA.  
AC P25624;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Topoisomerase II-associated protein PAT1.  
GN PAT1 OR YCR077C OR YCR77C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
OX NCBI\_TaxID=4932;  
RN [1] SEQUENCE FROM N.A.  
RP Ballesta J.P.G., Franco L., Hoenicka J., Jimenez A., Remacha M.,  
RA Sanz E.;  
RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.  
RN [2] REVISIONS, AND CHARACTERIZATION.  
RP MEDLINE=95373282; PubMed=7645349;  
RA Rodriguez-Cousino N., Lill R., Neupert W., Court D.A.;  
RT "Identification and initial characterization of the cytosolic protein  
Ycr77p.";  
RL Yeast 11:581-585(1995).  
RN [3] FUNCTION.  
RP MEDLINE=97128274; PubMed=8972867;  
RA Wang X., Watt P.M., Louis E.J., Borts R.H., Hickson I.D.;  
RT "Pat1: a topoisomerase II-associated protein required for faithful  
chromosome transmission in Saccharomyces cerevisiae.";  
RL Nucleic Acids Res. 24:4791-4797(1996).  
CC -!- FUNCTION. Necessary for accurate chromosome transmission during  
cell division. Interacts with topoisomerase TOP2.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -!- SIMILARITY: TO S.POMBE SPBC19G7.10C.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X59720; CAA42266.1; -;  
DR PIR; S53590; S53590.  
DR Germonline; 138577; -;  
DR SGD; S0000673; PAT1.  
DR GO; GO:0000932; C:cytoplasmic mRNA processing body; IDA.  
DR GO; GO:0005843; C:cytosolic small ribosomal subunit (sensu Eu. . .; IDA.  
DR GO; GO:0007059; P:chromosome segregation; IMP.  
DR GO; GO:0006446; P:regulation of translational initiation; IGI.  
FT DOMAIN 195 200 POLY-PRO.  
SQ SEQUENCE 797 AA; 88487 MW; 2A54F39AE3E75ECE CRC64;

Query Match 54.3%; Score 56.5; DB 1; Length 797;  
Best Local Similarity 62.8%; Pred. No. 12;  
Matches 11; Conservative 1; Mismatches 1; Indels 3; Gaps 1;

QY 2 GHRPPRGRCGPPQ 17  
DB 165 GH---PQGLPQPPQ 177

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:51:34 ; Search time 61.486 Seconds  
(without alignments)  
87.236 Million cell updates/sec

Title: US-10-009-709-13

Perfect score: 104

Sequence: 1 GGHPRPRGRPQGPPQQ 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriaph.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	68	65.4	173	4	Q15214
2	68	65.4	297	4	Q16038
3	68	65.4	338	4	Q861A1
4	68	65.4	358	4	Q08805
5	66.5	63.9	295	11	Q07611
6	64	61.5	188	6	Q29427
7	62	59.6	93	4	Q15215
8	62	59.6	238	4	Q00600
9	62	59.6	309	4	Q04118
10	62	59.6	382	4	Q00599
11	61	58.7	46	4	Q15218
12	59	56.7	950	11	Q923A9
13	59	56.7	2157	11	Q921R1
14	59	56.7	2158	11	Q7TSC1
15	58	55.8	170	11	Q07610
16	58	55.8	707	4	Q86VG2

```

17 58 55.8 715 5 Q9XW73
18 57 54.8 85 5 Q9VR11
19 57 54.8 352 5 Q9SRE4
20 57 54.8 359 5 Q46173
21 57 54.8 520 10 Q9LV14
22 56.5 54.3 796 3 Q8NKJ3
23 56 53.8 310 16 Q82EM1
24 56 53.8 504 11 Q1IX93
25 56 53.8 507 11 Q89038
26 56 53.8 507 11 Q92186
27 55.5 53.4 188 11 Q62106
28 55.5 53.4 317 11 Q62103
29 55 52.9 485 3 Q94115
30 54.5 52.4 300 11 Q61888
31 54.5 52.4 301 11 Q62105
32 54.5 52.4 333 4 Q96SD4
33 54.5 52.4 367 4 Q8N6Q2
34 54.5 52.4 367 4 Q81XT2
35 54 51.9 389 11 Q88539
36 54 51.9 407 4 Q96AY8
37 54 51.9 641 4 Q9Y2W2
38 54 51.9 641 11 Q223B5
39 54 51.9 641 11 Q8VD10
40 54 51.9 646 5 Q9U302
41 54 51.9 646 5 Q17350
42 53 51.0 96 11 Q8CD82
43 53 51.0 117 10 Q8H3N5
44 53 51.0 181 10 Q9ZRE9
45 53 51.0 254 10 Q8VZA7

```

## ALIGNMENTS

### RESULT 1

```

Q15214
ID Q15214 PRELIMINARY; PRT; 173 AA.
AC Q15214;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE Salivary proline-rich protein 1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini, Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84298176; PubMed=6089212;
RA Azen E.A., Lyons K.M., McGonigal T., Barrett N.L., Clements L.S.,
RA Maeda N., Vanin E.F., Carlson D.M., Smithies O.;
RT "Clones from the human gene complex coding for salivary proline-rich
RT proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:5561-5565 (1984).
DR EMBL; K02575; AAA36502.1;
DR GO; GO:0005576; C:extracellular; NAS.
FT NON_TER 1
FT NON_TER 173
SQ SEQUENCE 173 AA; 17206 MW; B84E30335B77A7C0 CRC64;

```

Query Match 65.4%; Score 68; DB 4; Length 173;  
Best Local Similarity 61.9%; Pred.No. 0.1;  
Matches 13; Conservative 2; Mismatches 2; Indels 4; Gaps 1;

```

QY 1 GGH----PRPRGRPQGPPQQ 17
    |||
    |||
Db 148 GGNQPGQPPPPGKPGQPPQQ 168

```

### RESULT 2

```

Q16038
ID Q16038 PRELIMINARY; PRT; 297 AA.
AC Q16038;

```

```

DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE PRB1M protein (Fragment).
GN PRB1M.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9330421; PubMed=8317492;
RA Azen E.A., Latreille P., Niece R.L.;
RT "PRB1 gene variants coding for length and null polymorphisms among
RT human salivary Ps, PmF, PmS, and Pe proline-rich proteins (PRPs).";
RL Am. J. Hum. Genet. 53:264-278(1993).
DR EMBL; S62928; AAB27288.2; -
DR PIR; B40750; PIHUB6.
DR PIR; D40750; D40750.
FT NON TER 1
SQ SEQUENCE 297 AA; 29046 MW; 1C7BFAC6B5B5F0 CRC64;

Query Match 65.4%; Score 68; DB 4; Length 297;
Best Local Similarity 61.9%; Pred. No. 0.17;
Matches 13; Conservative 2; Mismatches 2; Indels 4; Gaps 1;

QY 1 GGH----PRPRGRPGQPQQ 17
DB 131 GGNQPGPPPPGKPGQPQQ 151

RESULT 3
Q86YAL PRELIMINARY; PRT; 338 AA.
AC Q86YAL;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Skin;
RA Strausberg R.;
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC644827; AAH44827.1; -
KW Hypothetical protein.
FT NON TER 1
SQ SEQUENCE 338 AA; 33344 MW; 7F54B4E5AB002261 CRC64;

Query Match 65.4%; Score 68; DB 4; Length 338;
Best Local Similarity 61.9%; Pred. No. 0.2;
Matches 13; Conservative 2; Mismatches 2; Indels 4; Gaps 1;

QY 1 GGH----PRPRGRPGQPQQ 17
DB 172 GGNQPGPPPPGKPGQPQQ 192

RESULT 4
Q08805 PRELIMINARY; PRT; 358 AA.
AC Q08805;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE PRB1L protein (Fragment).
GN PRB1L.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9330421; PubMed=8317492;
RA Azen E.A., Latreille P., Niece R.L.;
RT "PRB1 gene variants coding for length and null polymorphisms among
RT human salivary Ps, PmF, PmS, and Pe proline-rich proteins (PRPs).";
RL Am. J. Hum. Genet. 53:264-278(1993).
DR EMBL; S62928; AAB27288.2; -
DR PIR; B40750; PIHUB6.
DR PIR; D40750; D40750.
FT NON TER 1
SQ SEQUENCE 358 AA; 35050 MW; DB7F87B8D5EA759E CRC64;

Query Match 65.4%; Score 68; DB 4; Length 358;
Best Local Similarity 61.9%; Pred. No. 0.21;
Matches 13; Conservative 2; Mismatches 2; Indels 4; Gaps 1;

QY 1 GGH----PRPRGRPGQPQQ 17
DB 192 GGNQPGPPPPGKPGQPQQ 212

RESULT 5
Q07611 PRELIMINARY; PRT; 295 AA.
AC Q07611;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Proline-rich proteoglycan.
GN PRPG2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Sprague-Dawley; TISSUE=Parotid gland;
RA Castle A.M., Castle J.D.;
RT "Novel secretory proline-rich proteoglycans from rat parotid. Cloning
RT and characterization by expression in AtT-20 cells.";
RL J. Biol. Chem. 268:20490-20496(1993).
DR EMBL; L17318; AAA03074.1; -
DR PIR; B48013; B48013.
SQ SEQUENCE 295 AA; 30026 MW; EEA44BD8B3E1B9F CRC64;

Query Match 63.9%; Score 66.5; DB 11; Length 295;
Best Local Similarity 72.2%; Pred. No. 0.27;
Matches 13; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

QY 1 GGHPRPPR-GRPGQPQQ 17
DB 264 GGPQRPQPGNPGQPQQ 281

RESULT 6
Q29427 PRELIMINARY; PRT; 188 AA.
AC Q29427;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Proline-rich protein.
GN MNP4 OR MNPBP9.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
```

```

RX MEDLINE=92009216; PubMed=1916292;
RA Lin H.H., Kousvelari E.E., Ann D.K.;
RT "Sequence and expression of the Mnp4 gene encoding basic proline-rich
RL protein in macaque salivary glands.";
RL Gene 104:219-226(1991).
DR EMBL; M81322; AAA36905.1; -.
DR EMBL; M81321; AAA36904.1; -.
DR PIR; JH0481; JH0481.
SQ SEQUENCE 188 AA; 19135 MW; C8E6D0B7F4DEB504 CRC64;

Query Match          61.5%; Score 64; DB 6; Length 188;
Best Local Similarity 63.2%; Pred. No. 0.38;
Matches 12; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

QY 1 GGHPR--PPRGRPQGPPQQ 17
   ||:|:|:|:|:|:|:|:|:|
DB 105 GNKPQGPFPFGKPGQGPQQ 123

RESULT 7
ID Q15215 PRELIMINARY; PRT; 93 AA.
AC Q15215;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Salivary proline-rich protein 1 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84298176; PubMed=6089212;
RA Azen E.A., Lyons K.M., McGonigal T., Barrett N.L., Clements L.S.,
RA Maeda N., Vanin E.F., Carlson D.M., Smithies O.;
RT "Clones from the human gene complex coding for salivary proline-rich
RT proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:5561-5565(1984).
DR EMBL; K02576; AAA36503.1; -.
FT NON TER 1
SQ SEQUENCE 93 AA; 9170 MW; 04AD5F0797E31867 CRC64;

Query Match          59.6%; Score 62; DB 6; Length 93;
Best Local Similarity 55.0%; Pred. No. 0.36;
Matches 11; Conservative 4; Mismatches 1; Indels 4; Gaps 1;

QY 1 GGHPR----PPRGRPQGPPQ 16
   ||:|:|:|:|:|:|:|:|:|
DB 62 GGNPQQPQAPPAGKPGGPPR 81

RESULT 8
ID Q00600 PRELIMINARY; PRT; 238 AA.
AC Q00600;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Farocid 'o' protein (Fragment).
GN PRB4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96108975; PubMed=8554050;
RA Azen E.A., Amberger E., Fisher S., Prakobphol A., Niece R.L.;
RT "PRB1, PRB2, and PRB4 coded polymorphisms among human salivary
RT concanavalin-A binding, II-1, and P0 proline-rich proteins.";
RL Am. J. Hum. Genet. 58:143-153(1996).
DR EMBL; S80916; AAB50897.2; -.

```

```

FT NON TER 1
SQ SEQUENCE 238 AA; 24005 MW; 4F4E947FFFA6C1 CRC64;

Query Match          59.6%; Score 62; DB 4; Length 238;
Best Local Similarity 57.9%; Pred. No. 0.89;
Matches 11; Conservative 3; Mismatches 1; Indels 4; Gaps 1;

QY 1 GGHPR----PPRGRPQGPP 15
   ||:|:|:|:|:|:|:|:|:|
DB 201 GGNPQQPQAPPAGKPGGPP 219

RESULT 9
ID Q04118 PRELIMINARY; PRT; 309 AA.
AC Q04118;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Proline-rich protein G1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89121440; PubMed=2851479;
RA Lyons K.M., Stein J.H., Smithies O.;
RT "Length polymorphisms in human proline-rich protein genes generated by
RT intragenic unequal crossing over.";
RL Genetics 120:267-278(1998).
DR EMBL; X07881; CAA30728.1; -.
DR PIR; S10889; S10889.
DR Genew; HGNC:9339; PRB3.
SQ SEQUENCE 309 AA; 30936 MW; 2C14D616974320C8 CRC64;

Query Match          59.6%; Score 62; DB 4; Length 309;
Best Local Similarity 57.9%; Pred. No. 1.1;
Matches 11; Conservative 3; Mismatches 1; Indels 4; Gaps 1;

QY 1 GGHPR----PPRGRPQGPP 15
   ||:|:|:|:|:|:|:|:|:|
DB 273 GGNPQQPLPPAGKPGGPP 291

RESULT 10
ID O00599 PRELIMINARY; PRT; 382 AA.
AC O00599;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Con1 (Fragment).
GN PRB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96108975; PubMed=8554050;
RA Azen E.A., Amberger E., Fisher S., Prakobphol A., Niece R.L.;
RT "PRB1, PRB2, and PRB4 coded polymorphisms among human salivary
RT concanavalin-A binding, II-1, and P0 proline-rich proteins.";
RL Am. J. Hum. Genet. 58:143-153(1996).
DR EMBL; S80905; AAB50686.1; -.
FT NON TER 1
SQ SEQUENCE 382 AA; 37277 MW; D500942AB979209D CRC64;

Query Match          59.6%; Score 62; DB 4; Length 382;
Best Local Similarity 55.0%; Pred. No. 1.4;
Matches 11; Conservative 4; Mismatches 1; Indels 4; Gaps 1;

```

```

QY 1 GGHPR----PPRGRPGQPP 16
Db 351 GGNPQPPQAPPQPP 370

RESULT 11
Q15218 PRELIMINARY; PRT; 46 AA.
AC Q15218;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DE Salivary proline-rich protein 2 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini, Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84298176; PubMed=6089212;
RA Azen E.A., Lyons K.M., McGonigal T., Barrett N.L., Clements L.S.,
RA Maeda N., Vanin E.F., Carlson D.M., Smithies O.;
RT "Clones from the human gene complex coding for salivary proline-rich
RT proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 81:5561-5565 (1984).
DR EMBL; K02578; AAA36505.1; -.
FT NON_TER 1
FT NON_TER 46
SQ SEQUENCE 46 AA; 4592 MW; FCE1D38D8DEDC173 CRC64;

Query Match 58.7%; Score 61; DB 4; Length 46;
Best Local Similarity 57.9%; Pred. No. 0.25;
Matches 11; Conservative 3; Mismatches 1; Indels 4; Gaps 1;

QY 1 GGHPR----PPRGRPGQPP 15
Db 28 GGNPQPPQAPPQPP 46

RESULT 12
Q923A9 PRELIMINARY; PRT; 950 AA.
AC Q923A9;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Salivary proline-rich protein 2 (Fragment).
DE Similar to RIKEN cDNA 3110039B05 gene (Fragment).
GN BAT2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC006664; AAH06664.1; -.
FT MGD; MGI:1915467; Bat2.
FT NON_TER 1
SQ SEQUENCE 950 AA; 99726 MW; E41DE9AB4A6D976 CRC64;

Query Match 56.7%; Score 59; DB 11; Length 950;
Best Local Similarity 71.4%; Pred. No. 8.4;
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GGHPRPPRGRPGQPP 14
Db 282 GGHPRPKPGPPQAP 295

RESULT 13
Q921R1 PRELIMINARY; PRT; 2157 AA.
ID Q921R1

```

```

AC Q921R1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE BAT2.
DE GN NFKB1L1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowen L., Qin S., Madan A., Abbasi N., James R., Dickhoff R.,
RA Shaffer T., Ratcliffe A., Loretz C., Lasky S., Hood L.;
RT "Sequence of the mouse major histocompatibility class III region.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF109719; AAC82480.1; -.
DR MGD; MGI:1340031; NFKB1L1.
SQ SEQUENCE 2157 AA; 229072 MW; 35B21F6B7C971F40 CRC64;

Query Match 56.7%; Score 59; DB 11; Length 2157;
Best Local Similarity 71.4%; Pred. No. 18;
Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GGHPRPPRGRPGQPP 14
Db 1490 GGHPRPKPGPPQAP 1503

RESULT 14
Q7TSC1 PRELIMINARY; PRT; 2158 AA.
ID Q7TSC1
AC Q7TSC1;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DE Bat2 protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RX MEDLINE=23188257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahay J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Krzywinski M.I., Skalska U., Smalus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strausberg R.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC053522; AAH53522.1; -.
SQ SEQUENCE 2158 AA; 229201 MW; 85A5CBDCDA84A256 CRC64;

```

Query Match 56.7%; Score 59; DB 11; Length 2158;  
 Best Local Similarity 71.4%; Pred. No. 18;  
 Matches 10; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 GGHPRPRGRPGQP 14  
 |||||  
 Db 1490 GGHPRPKPGPPQAP 1503

## RESULT 15

Q07610  
 ID Q07610 PRELIMINARY; PRT; 170 AA.  
 AC Q07610;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Proline-rich proteoglycan.  
 GN PRG1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Parotid gland;  
 RX MEDLINE=93388626; PubMed=8376404;  
 RA Castle A.M., Castle J.D.;  
 RT "Novel secretory proline-rich proteoglycans from rat parotid. Cloning  
 RL and characterization by expression in AtT-20 cells.";  
 RL J. Biol. Chem. 268:20490-20496(1993).  
 DR EMBL; L17317; AAA03073.1; -.  
 DR FTR; A48013; A48013.  
 SQ SEQUENCE 170 AA; 17840 MW; 85AF34C40B45AA03 CRC64;

Query Match 55.8%; Score 58; DB 11; Length 170;  
 Best Local Similarity 64.7%; Pred. No. 2.2;  
 Matches 11; Conservative 2; Mismatches 2; Indels 2; Gaps 1;

QY 2 GHPR--PPGRPGGPPQ 16  
 |||:|||||  
 Db 111 GKPQPPPGGPPQ 127

Search completed: April 6, 2004, 16:14:44  
 Job time : 61.5629 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:29:39 ; Search time 95.1682 Seconds  
(without alignments)  
50.472 Million cell updates/sec

Title: US-10-009-709-13

Perfect score: 104

Sequence: 1 GCHPRPRGRPGGPQQ 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1596107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_29Jan04.\*

- 1: Geneseq1980s.\*
- 2: Geneseq1990s.\*
- 3: Geneseq2000s.\*
- 4: Geneseq2001s.\*
- 5: Geneseq2002s.\*
- 6: Geneseq2003as.\*
- 7: Geneseq2003bs.\*
- 8: Geneseq2004s.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	17	AA030407	Human sal
2	104	100.0	132	AA030246	Human Sap
3	104	100.0	132	AA030188	Human cp3
4	104	100.0	132	ABU11891	Human ABC
5	104	100.0	132	AA030408	Human sec
6	104	100.0	132	ABG15767	Novel hum
7	104	100.0	132	ABG18110	Novel hum
8	104	100.0	132	AAW03556	Octadecap
9	104	100.0	154	AA022606	Zea mays
10	104	100.0	166	AA000989	Human pol
11	104	100.0	166	AA000989	Human pol
12	70	67.3	106	AA000989	Human sal
13	70	67.3	106	AA000989	Human sal
14	69	66.3	44	ABP76681	Streptomy
15	69	66.3	45	ABP76681	Streptomy
16	68	65.4	331	ABP76681	Streptomy
17	66.5	63.9	296	ABP76681	Streptomy
18	66.5	63.9	296	ABP76681	Streptomy
19	64	61.5	111	ABP76681	Streptomy
20	64	61.5	111	ABP76681	Streptomy
21	62	59.6	10	ABP76681	Streptomy
22	62	59.6	49	ABP76681	Streptomy
23	62	59.6	82	ABP76681	Streptomy
24	62	59.6	117	ABP76681	Streptomy
25	62	59.6	124	ABP76681	Streptomy

26	62	59.6	162	7	AA030407	Human sec
27	62	59.6	132	6	AA030246	Human Sap
28	62	59.6	208	6	AA030188	Human cp3
29	62	59.6	247	6	ABU11891	Human ABC
30	62	59.6	267	7	AA030408	Human sec
31	62	59.6	591	4	ABG15767	Novel hum
32	62	59.6	591	4	ABG18110	Novel hum
33	61	58.7	18	2	AAW03556	Octadecap
34	59.5	57.2	158	3	AA022606	Zea mays
35	59	56.7	125	4	AA000989	Human pol
36	58	55.8	10	4	AA000989	Human pol
37	58	55.8	19938	6	ABP76681	Streptomy
38	57	54.8	85	4	ABP76681	Streptomy
39	57	54.8	91	3	AA034161	Gene 17 h
40	57	54.8	315	4	ABG1702	Drosophil
41	56.5	54.3	106	3	AA026069	Zea mays
42	56.5	54.3	796	6	ABG53413	Protein s
43	56	53.8	9	4	AA048778	Human sal
44	56	53.8	41	6	AA030248	Human Sap
45	56	53.8	64	3	AA05869	Germline

#### ALIGNMENTS

RESULT 1  
AAB48783  
ID AAB48783 standard; peptide; 17 AA.  
XX  
AC AAB48783;  
XX  
DT 09-MAR-2001 (first entry)  
XX  
DE Human saliva PRP-1 fragment (residues 99-115), SEQ ID NO:13.  
XX  
KW Human PRP-1; proline-rich protein; saliva; dental caries;  
KW chromosome l2p13.2; arginine catabolism; ammonia production; pH increase;  
KW oral bacterium; caries prevention.  
XX  
OS Homo sapiens.  
XX  
PN WO200069890-A1.  
XX  
PD 23-NOV-2000.  
XX  
PF 11-MAY-2000; 2000WO-SE000930.  
XX  
PR 17-MAY-1999; 99SE-00001773.  
XX  
PA (STRO/) STROEMBERG N.  
PA (JOHA/) JOHANSSON I.  
XX  
PI Stroemberg N, Johansson I;  
XX  
DR WPI; 2001-031923/04.  
XX  
PT New oligopeptides comprising 2 arginine residues from degradation of  
PT proline-rich proteins, useful for preventing dental caries.  
XX  
PS Claim 2; Page 24; 36pp; English.  
XX  
CC The invention relates to human PRP-1-derived oligopeptides (AAB48771-  
CC AAB48783) which contain at least two arginine residues and which protect  
CC against dental caries. PRPs (proline-rich proteins) are salivary proteins  
CC encoded by six clustered genes on chromosome l2p13.2 and are potential  
CC determinants of a person's susceptibility to dental caries. PRPs are  
CC degraded by Actinomyces and Streptococcus species to small peptide  
CC fragments. These are metabolised by oral bacteria for nutritional  
CC purposes, with certain bacterial species generating ammonia via the  
CC catabolism of arginine. The peptides of the invention, being arginine-  
CC rich, can also be converted to ammonia by these bacteria. The ammonia  
CC thus formed raises the pH at the dental surface, thereby protecting the  
CC teeth against caries. Sequences AAB48771-B48783 represent the PRP-1-



CC derived oligopeptides of the invention  
XX  
SQ Sequence 17 AA;

Query Match 100.0%; Score 104; DB 4; Length 17;  
Best Local Similarity 100.0%; Pred. No. 6.6e-06;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRPRGRPGPPQQ 17  
Db 1 GGHPRPRGRPGPPQQ 17

## RESULT 2

ABB38848  
ID ABB38848 standard; peptide; 132 AA.

XX  
AC ABB38848;

XX  
DT 04-FEB-2002 (first entry)

XX Peptide #6354 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

XX WO200157277-A2.

XX  
XX 09-AUG-2001.

XX  
XX 30-JAN-2001; 2001WO-US000669.

XX  
XX 04-FEB-2000; 2000US-0180312P.

XX  
XX 26-MAY-2000; 2000US-0207456P.

XX  
XX 30-JUN-2000; 2000US-00608408.

XX  
XX 03-AUG-2000; 2000US-00632366.

XX  
XX 21-SEP-2000; 2000US-0234687P.

XX  
XX 27-SEP-2000; 2000US-0236359P.

XX  
XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human foetal liver.

XX Claim 27; SEQ ID NO 31483; 639pp + Sequence Listing; English.

XX The invention relates to a single exon nucleic acid probe for measuring

XX human gene expression in a sample derived from human foetal liver. The

XX single exon nucleic acid probes may be used for predicting, measuring and

XX displaying gene expression in samples derived from human foetal liver. The

XX present sequence is a peptide encoded by a single exon nucleic acid probe

XX of the invention. Note: The sequence data for this patent did not form

XX part of the printed specification, but was obtained in electronic format

XX directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 104; DB 4; Length 132;

Best Local Similarity 100.0%; Pred. No. 4.1e-05;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRPRGRPGPPQQ 17  
Db 81 GGHPRPRGRPGPPQQ 97

## RESULT 3

AAM32323  
ID AAM32323 standard; protein; 132 AA.

XX  
XX AAM32323;

XX  
XX 17-OCT-2001 (first entry)

XX Peptide #6360 encoded by probe for measuring placental gene expression.

XX Probe; microarray; human; placenta; antenatal diagnosis;

XX genetic disorder.

XX Homo sapiens.

XX WO200157272-A2.

XX  
XX 09-AUG-2001.

XX  
XX 30-JAN-2001; 2001WO-US000663.

XX  
XX 04-FEB-2000; 2000US-0180312P.

XX  
XX 26-MAY-2000; 2000US-0207456P.

XX  
XX 30-JUN-2000; 2000US-00608408.

XX  
XX 03-AUG-2000; 2000US-00632366.

XX  
XX 21-SEP-2000; 2000US-0234687P.

XX  
XX 27-SEP-2000; 2000US-0236359P.

XX  
XX 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488897/53.

XX Human genome-derived single exon nucleic acid probes useful for analyzing

XX gene expression in human placenta.

XX Claim 27; SEQ ID NO 32592; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP:

XX see AAI31315-AI57546). The present sequence is a peptide encoded by one

XX such probe. The probes are useful for producing a microarray for

XX predicting, measuring and displaying gene expression in samples derived

XX from human placenta. The probes are useful for antenatal diagnosis of

XX human genetic disorders

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 104; DB 4; Length 132;

Best Local Similarity 100.0%; Pred. No. 4.1e-05;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRPRGRPGPPQQ 17  
Db 81 GGHPRPRGRPGPPQQ 97

## RESULT 4

AAM72058

ID AAM72058 standard; protein; 132 AA.

XX  
XX AAM72058;

XX  
XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed probe encoded protein SEQ ID NO: 32364.

XX Human; bone marrow expressed exon; gene expression analysis; probe;

XX microarray; cancer; leukaemia; lymphoma; myeloma.

XX Homo sapiens.

XX WO200157276-A2.

```

XX PD 09-AUG-2001.
XX PF
XX 30-JAN-2001; 2001WO-US000668.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX DR
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human bone marrow.
XX PS Example 4; SEQ ID NO 32364; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX bone marrow. They can be used to measure gene expression in bone marrow
XX samples, which may enable the improved diagnosis and treatment of cancers
XX such as lymphoma, leukaemia and myeloma. The present sequence is a
XX protein encoded by one of the probes of the invention
XX SQ Sequence 132 AA;

Query Match 100.0%; Score 104; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 4.1e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRPRGRPGQPQQ 17
DB 81 GGHPRPRGRPGQPQQ 97

RESULT 5
AAM59494
ID AAM59494 standard; protein; 132 AA.
XX AC
XX AAM59494;
XX DT 05-NOV-2001 (first entry)
XX DE Human brain expressed single exon probe encoded protein SEQ ID NO: 31599.
XX KW Human; brain expressed exon; gene expression analysis; probe; microarray;
XX KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
XX OS Homo sapiens.
XX PN WO200157275-A2.
XX PD
XX PF
XX 30-JAN-2001; 2001WO-US000667.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX DR
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX PS Claim 27; SEQ ID NO 32390; 658pp; English.
XX CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridises at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.

```

```

PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483446/52.
XX PT Single exon nucleic acid probes for analyzing gene expression in human
XX brains.
XX PS Example 4; SEQ ID NO 31599; 650pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX probes which are derived from genomic sequences expressed in the human
XX brain. They can be used to measure gene expression in brain cell samples,
XX which may enable the diagnosis and improved treatment of nervous system
XX diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
XX epilepsy and cancers. The present sequence is a protein encoded by one of
XX the probes of the invention
XX SQ Sequence 132 AA;

Query Match 100.0%; Score 104; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 4.1e-05;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRPRGRPGQPQQ 17
DB 81 GGHPRPRGRPGQPQQ 97

RESULT 6
ABG53742
ID ABG53742 standard; peptide; 132 AA.
XX AC
XX ABG53742;
XX DT 25-FEB-2003 (first entry)
XX DE Human liver peptide, SEQ ID NO 32390.
XX KW Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
XX KW hypercholesterolaemia; coronary heart disease.
XX OS Homo sapiens.
XX PN WO200157273-A2.
XX PD
XX PF
XX 30-JAN-2001; 2001WO-US000664.
XX PR 04-FEB-2000; 2000US-0180312P.
XX PR 26-MAY-2000; 2000US-0207456P.
XX PR 30-JUN-2000; 2000US-00608408.
XX PR 03-AUG-2000; 2000US-00632366.
XX PR 21-SEP-2000; 2000US-0234687P.
XX PR 27-SEP-2000; 2000US-0236359P.
XX PR 04-OCT-2000; 2000GB-00024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488998/53.
XX DR
XX PT Human genome-derived single exon nucleic acid probes useful for analyzing
XX gene expression in human adult liver.
XX PS Claim 27; SEQ ID NO 32390; 658pp; English.
XX CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
XX measuring human gene expression in a sample derived from human adult
XX liver, comprising one of 13109 defined nucleotide sequences given in the
XX specification (or complements/ fragments). The probe hybridises at high
XX stringency to a nucleic acid molecule expressed in the human adult liver.

```

CC (I) may be used for predicting, measuring and displaying gene expression  
 CC in samples derived from human adult liver. The genes identified may be  
 CC involved in genetic liver diseases such as cirrhosis,  
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is  
 CC associated with coronary heart disease. ABG47348-ABG5930 represent human  
 CC liver single exon encoded peptides of the invention. Note: The sequence  
 CC information for this patent does not appear in the printed specification  
 CC but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 104; DB 4; Length 132;

Best Local Similarity 100.0%; Pred. No. 4.1e-05;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRRPRGRPGPPQQ 17

Db 81 GGHPRRPRGRPGPPQQ 97

RESULT 7

ABG41873

ID ABG41873 standard; peptide; 132 AA.

AC ABG41873;

DT 19-AUG-2002 (first entry)

DE Human peptide encoded by genome-derived single exon probe SEQ ID 31538.

KW Human; single exon probe; asthma; lung cancer; COPD; ILD;

KW chronic obstructive pulmonary disease; interstitial lung disease;

KW familial idiopathic pulmonary fibrosis; neurofibromatosis;

KW tuberous sclerosis; Gaucher's disease; Niemann-Pick disease;

KW Hermansky-Pudlak syndrome; sarcoidosis; pulmonary haemosiderosis;

KW pulmonary histiocytosis; lymphangioma; fibrocystic pulmonary

KW primary alveolar proteinosis; fibrocystic pulmonary dysplasia;

KW primary ciliary dyskinesia; pulmonary hypertension;

KW hyaline membrane disease.

XX Homo sapiens.

OS WO2001:86003-A2.

PN 15-NOV-2001.

PD 30-JAN-2001; 2001WO-US000665.

PF 04-FEB-2000; 2000US-0180312P.

PR 26-MAY-2000; 2000US-0207456P.

PR 30-JUN-2000; 2000US-00608408.

PR 03-AUG-2000; 2000US-00632366.

PR 21-SEP-2000; 2000US-0234687P.

PR 27-SEP-2000; 2000US-0236359P.

PR 04-OCT-2000; 2000GB-00024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

PA Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2002-114183/15.

XX Spatially-addressable set of single exon nucleic acid probes, used to

PT measure gene expression in human lung samples.

PT Claim 27; SEQ ID NO 31538; 634pp; English.

XX The invention relates to a spatially-addressable set of single exon

XX nucleic acid probes for measuring gene expression in a sample derived

XX from human lung comprising single exon nucleic acid probes having one of

XX 12614 nucleic acid sequences mentioned in the specification, or their

XX complements or the 12387 open reading frames derived from the 12614

CC probes. Also included are a microarray comprising the novel set of probes  
 CC ; the novel set of probes which hybridise at high stringency to a nucleic  
 CC acid expressed in the human lung; measuring gene expression in a sample  
 CC derived from human lung, comprising (a) contacting the array with a  
 CC collection of detectably labeled nucleic acids derived from human lung  
 CC mRNA, and (b) measuring the label detectably bound to each probe of the  
 CC array; identifying exons in a eukaryotic genome, comprising (a)  
 CC algorithmically predicting at least one exon from genomic sequences of  
 CC the eukaryote; and (b) detecting specific hybridisation of detectably  
 CC labeled nucleic acids from eukaryote lung mRNA, to a single exon probe,  
 CC having a fragment identical to the predicted exon, the probe is included  
 CC in the above mentioned microarray; assigning exons to a single gene,  
 CC comprising (a) identifying exons from genomic sequence by the method  
 CC above and (b) measuring the expression of each of the exons in several  
 CC tissues and/or cell types using hybridisation to a single exon  
 CC microarrays having a probe with the exon, where a common pattern of  
 CC expression of the exons in the tissues and/or cell types indicates that  
 CC the exons should be assigned to a single gene; a peptide comprising one  
 CC of 12011 sequences, mentioned in the specification, or encoded by the  
 CC probes/open reading frames (ORF). The probes are used for gene expression  
 CC analysis, and for identifying exons in a gene, particularly using human  
 CC lung derived mRNA and for the study of lung diseases such as asthma, lung  
 CC cancer, chronic obstructive pulmonary disease (COPD), interstitial lung  
 CC disease (ILD), familial idiopathic pulmonary fibrosis, neurofibromatosis,  
 CC tuberous sclerosis, Gaucher's disease, Niemann-Pick disease, Hermansky-  
 CC Pudlak syndrome, sarcoidosis, pulmonary haemosiderosis, pulmonary  
 CC histiocytosis, lymphangioma, fibrocystic pulmonary alveolar proteinosis,  
 CC Karagener syndrome, fibrocystic pulmonary dysplasia, primary ciliary  
 CC dyskinesia, pulmonary hypertension and hyaline membrane disease. The  
 CC present sequence is a peptide/protein encoded by a single exon probe of  
 CC the invention. Note: The sequence data for this patent did not form part  
 CC of the printed specification, but was obtained in electronic format  
 CC directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX SQ Sequence 132 AA;

Query Match 100.0%; Score 104; DB 5; Length 132;

Best Local Similarity 100.0%; Pred. No. 4.1e-05;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GGHPRRPRGRPGPPQQ 17

Db 81 GGHPRRPRGRPGPPQQ 97

RESULT 8

ABR57423

ID ABR57423 standard; protein; 149 AA.

AC ABR57423;

DT 15-SEP-2003 (first entry)

XX Human NOV7 protein SEQ ID NO:24.

DE Human; NOV7; cytostatic; cardiant; antiinflammatory; immunosuppressive;

XX anti-allergic; haemostatic; anti-HIV; antidiabetic; antiarteriosclerotic;

XX anorectic; antiaesthetic; nephrotropic; antiaarthritis; hepatotropic;

XX neuroprotective; nootropic; antibacterial; virucide; antiparasitic;

XX relaxant; anticonvulsant; hypotensive; vasotropic; antiparkinsonian;

XX vulnary; angiogenic; antiangiogenic; gene therapy; vaccine; cancer;

XX cardiomyopathy; atherosclerosis; hypertension; diabetes; inflammation;

XX autoimmune disorder; allergy; blood disorder; AIDS; obesity; asthma;

XX acquired immunodeficiency syndrome; nephropathy; cirrhosis; arthritis;

XX Alzheimer's disease; Parkinson's disease; goitre; infection; stroke;

XX muscular dystrophy; epilepsy; wasting disorder; chromosome 12.

XX Homo sapiens.

OS WO200294870-A2.

XX 28-NOV-2002.

XX

02-NOV-2001; 2001WO-US051580.  
 02-NOV-2000; 2000US-0245291P.  
 02-NOV-2000; 2000US-0245317P.  
 07-NOV-2000; 2000US-0246562P.  
 08-NOV-2000; 2000US-0246871P.  
 26-JAN-2001; 2001US-0264389P.  
 26-JAN-2001; 2001US-0264423P.  
 29-JAN-2001; 2001US-0264799P.  
 (CURA-) CURAGEN CORP.  
 Grosse WM, Macdougall JR, Smithson G, Millet I, Stone DJ;  
 Gunther E, Ellerman K, Alsobrook JP, Lepley DM, Burgess CE;  
 Spytek KA, Edinger SR, Gargolli EA, Gorman L, Taupier RJ, Li L;  
 Guo X, Fernandes ER, Vernet CAM, Tchernev VT, Casman SJ, Shenoy S;  
 Mishra V, Furtak K, Baumgartner JC, Colman SD;  
 WPI; 2003-140359/13.  
 N-PSDB; ACF03558.  
 New NOVX polypeptide useful for preventing or treating NOVX-associated  
 disorders, e.g. cancer, cardiomyopathy, atherosclerosis or diabetes, and  
 in chromosome mapping, tissue typing or pharmacogenomics.  
 Claim 1; Page 69; 346pp; English.  
 ACF03547 to ACF03570 encode the human NOVX proteins (I) given in ABR57412  
 to ABR57435. (I) have cytostatic, cardiant, antiinflammatory, nootropic,  
 immunosuppressive, antiallergic, haemostatic, anti-HIV, antidiabetic,  
 antiarteriosclerotic, anorectic, antiasthmatic, nephrotropic, virucide,  
 antiarthritic, hepatotropic, neuroprotective, antibacterial, relaxant,  
 antiparasitic, anticonvulsant, hypotensive, vasotropic, antiparkinsonian,  
 vulnary, angiogenic and antiangiogenic activities, and can be used in  
 gene therapy and vaccines. The NOVX polypeptides and their antibodies can  
 be used to determine the presence or absence of (I) in a sample. The NOVX  
 polypeptides, polynucleotides encoding them, and antibodies against them,  
 are useful in manufacturing a medicament for treating or preventing a  
 syndrome associated with a NOVX-associated disorder such as hypertension,  
 cardiomyopathy, atherosclerosis, cancer, diabetes, asthma, inflammation,  
 autoimmune disorders, allergies, blood disorders, obesity, acquired  
 immunodeficiency syndrome (AIDS), immunoglobulin (Ig) A nephropathy,  
 cirrhosis, arthritis, Alzheimer's disease, Parkinson's disease, goitre,  
 infections (e.g. bacterial, viral, parasitic), stroke, muscular  
 dystrophy, epilepsy, and other wasting disorders associated with chronic  
 diseases. ACF03571 to ACF03644 represent PCR primers and probes for NOVX  
 sequence, which are used in an example from the present invention  
 Sequence 149 AA;  
 Query Match 100.0%; Score 104; DB 6; Length 149;  
 Best Local Similarity 100.0%; Pred. No. 4.5e-05;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGHPRRPRGRPGQPQQ 17  
 Db 98 GGHPRRPRGRPGQPQQ 114  
 RESULT 9  
 ABR56769  
 ID ABR56769 standard; protein; 154 AA.  
 AC ABR56769;  
 XX  
 DT 30-JUL-2003 (first entry)  
 XX  
 DE Human secreted protein SECP-44 SEQ ID NO:44.  
 XX  
 KW Human; secreted protein; SECP; cytostatic; antiarteriosclerotic;  
 KW anticonvulsant; nootropic; neuroprotective; cerebroprotective; anti-HIV;  
 KW antiallergic; antiinflammatory; thyromimetic; gene therapy; cancer;  
 KW cell proliferative disorder; atherosclerosis; neurological disorder;

KW epilepsy; Huntington's disease; stroke; immune disorder; allergy; AIDS;  
 KW inflammatory disorder; developmental disorder; hypothyroidism;  
 XX Cushing's syndrome; infection.  
 OS Homo sapiens.  
 XX WO2003016506-A2.  
 PN 27-FEB-2003.  
 PD 15-AUG-2002; 2002WO-US027143.  
 XX 17-AUG-2001; 2001US-0313249P.  
 XX 24-AUG-2001; 2001US-0314752P.  
 PR 07-SEP-2001; 2001US-0317818P.  
 PR 07-SEP-2001; 2001US-0317824P.  
 PR 21-SEP-2001; 2001US-0324040P.  
 PR 24-SEP-2001; 2001US-0324586P.  
 PR 02-NOV-2001; 2001US-0343980P.  
 PR 28-NOV-2001; 2001US-0334229P.  
 PR 13-FEB-2002; 2002US-0357002P.  
 PR 06-MAR-2002; 2002US-0362439P.  
 PR 19-MAR-2002; 2002US-0366041P.  
 PR 30-APR-2002; 2002US-0376988P.  
 XX (INCY-) INCYTE GENOMICS INC.  
 XX Tang TY, Warren BA, Gietzen KJ, Lal PG, Yue H, Honchell CD;  
 Lehr-Mason PM, Burford N, Xu Y, Baughn MR, Duggan BM, Iran UK;  
 PI Lee EA, Forsythe IJ, Richardson TW, Lee S, Thangavelu K, Yue H;  
 PI Emerling BM, Walla NK, Azimzai Y, Sanjanwala B, Rafalia AJA;  
 PI Borowsky ML, Nguyen DB, Tson CH, Aströmoff A, Ding L, Lee SY;  
 PI Becha SD, Ramkumar J, Gandhi AR, Jin P, Fu GK, Swarnakar A;  
 XX WPI; 2003-278569/27.  
 DR N-PSDB; ACC79069.  
 DR New human secreted proteins (SECP), useful for diagnosing, treating and  
 PT preventing diseases or conditions associated with the aberrant SECP  
 PT expression e.g. cancer, AIDS, atherosclerosis, Huntington's disease,  
 PT stroke, infections.  
 XX Claim 1; Page 222; 286pp; English.  
 PS ACC79026 to ACC79105 encode the human secreted proteins (I) given in  
 CC ABR56726 to ABR56805, designated SECP-1 to SECP-80. SECP sequences can  
 CC have cytostatic, antiarteriosclerotic, anticonvulsant, antiinflammatory,  
 CC neurotropic, neuroprotective, cerebroprotective, anti-HIV, antiallergic and  
 CC thyromimetic activities, and can be used in gene therapy. The SECP  
 CC preventing diseases or conditions associated with the decreased  
 CC expression or overexpression of SECP, such as cell proliferative (e.g.  
 CC cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's  
 CC disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and  
 CC developmental (e.g. hypothyroidism, Cushing's syndrome) disorders, or  
 CC infections. They are also useful in assessing the effects of exogenous  
 CC compounds on the expression of nucleic acid and amino acid sequences of  
 CC SECP. The SECP or its fragments are useful in screening compounds for  
 CC effectiveness as agonist or antagonist of the polypeptides, or in  
 CC altering the expression of the target polynucleotide and compounds that  
 CC specifically bind to or modulate the activity of the polypeptide  
 XX  
 XX Sequence 154 AA;  
 Query Match 100.0%; Score 104; DB 6; Length 154;  
 Best Local Similarity 100.0%; Pred. No. 4.7e-05;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GGHPRRPRGRPGQPQQ 17  
 Db 103 GGHPRRPRGRPGQPQQ 119

## RESULT 10

ADA83798  
 ID ADA83798 standard; protein; 166 AA.

XX AC ADA83798;

XX DT 20-NOV-2003 (first entry)

XX DE Human PRH2 protein.

XX KW human; marker; expressed sequence tag; EST; arabinidopsis; tumour;  
 KW stress-induced phenotype; hyperosmotic stress; colon cancer; immunogen;  
 KW vaccine.

XX OS Homo sapiens.

XX PN WO2002103028-A2.

XX PD 27-DEC-2002.

XX PF 30-MAY-2002; 2002WO-IB004189.

XX PR 22-MAY-2001; 2001US-0293999P.

XX PR 22-OCT-2001; 2001US-0330457P.

XX PR 19-FEB-2002; 2002US-0357144P.

XX PA (BIOM-) BIOMEDICAL CENT.

XX PI Baranova AV, Yankovsky NK, Kozlov AP, Lobashev AV, Krukovskaya LL;

XX DR MPI; 2003-175241/17.

XX DR N-PSDB; ADA83797.

XX PT Determining if a nucleic acid is a marker for a phenotype/cell type of  
 PT interest, by global comparison of expressed sequence tags known to be  
 PT expressed in the phenotype/cell type with all ESTs expressed in normal  
 PT tissue.

XX PS Claim 29; Page 191-192; 516pp; English.

XX CC The invention relates to a novel method for determining if a nucleic acid  
 CC is a marker for a predetermined phenotype/cell type of interest from a  
 CC biological species. The method comprises performing a global comparison  
 CC of a group of expressed sequence tags (ESTs) known to be expressed in the  
 CC phenotype/cell type of interest with all ESTs expressed in normal tissue  
 CC in order to identify ESTs that are preferentially expressed in the  
 CC phenotype/cell type of interest. A method of the invention is useful for  
 CC determining whether a nucleic acid is a marker for a predetermined  
 CC phenotype or cell type of interest from a biological species, preferably  
 CC Arabidopsis or human. The cell type of interest is an abnormal cell such  
 CC as a tumour cell, and the predetermined phenotype is a stress-induced  
 CC phenotype such as hyperosmotic stress or high salt conditions. A method  
 CC of the invention is also useful for determining the progression of colon  
 CC cancer in a human, for detecting a tumour cell, and for regulating or  
 CC preventing the growth of a tumour cell. An antibody of the invention is  
 CC useful for detecting the absence or presence of peptides encoded by  
 CC tumour-associated markers. A polypeptide of the invention is useful as an  
 CC immunogen for vaccinating an animal. The present sequence represents a  
 CC tumour-associated antigen of the invention.

XX SQ Sequence 166 AA;

Query Match

Best Local Similarity 100.0%; Score 104; DB 6; Length 166;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGHPRPGRGPPQQ 17

Db 115 GGHPRPGRGPPQQ 131

## RESULT 11

ADA8216

ADC98216 standard; protein; 166 AA.

XX AC ADC98216;

XX DT 01-JAN-2004 (first entry)

XX DE Human salivary acidic proline-rich phosphoprotein (PRP).

XX KW Human; salivary acidic proline-rich phosphoprotein; PRP; autoantigen;  
 KW atopic dermatitis-inducing protein; salivary gland; Igs autoantibody;  
 KW immunoglobulin E; mast cell activation; basophil activation; diagnosis;  
 KW risk assessment; sensitisation remedy; dermatological; antiallergic;  
 KW antiinflammatory.

XX OS Homo sapiens.

XX PN WO2003094991-A1.

XX PD 16-OCT-2003.

XX PF 04-APR-2003; 2003WO-JP004325.

XX PR 08-APR-2002; 2002JP-00105425.

XX PA (NISC-) JAPAN SCI & TECHNOLOGY CORP.

XX PI Hide M, Yamamoto S, Tanaka T, Koro O;

XX DR MPI; 2003-833567/77.

XX DR N-PSDB; ADC98215.

XX PT Atopic dermatitis-inducing proteins, applicable in diagnosis of including  
 PT risk of onset, and in developing sensitization remedies for the disease.

XX PS Claim 4; SEQ ID NO 2; 43pp; Japanese.

XX CC The invention relates to the human atopic dermatitis-inducing proteins,  
 CC salivary acidic proline-rich phosphoprotein (PRP; ADC98216) and prolactin  
 CC -inducible protein (PIP; ADC98218), and their post-translationally  
 CC modified forms. These proteins are secreted by salivary or sweat glands  
 CC and bind to Igs autoantibodies, thereby activating mast cells and  
 CC basophils. The invention also relates to antigenic peptide fragments of  
 CC PRP or PIP; antibodies which bind to PRP or PIP; methods for diagnosing  
 CC atopic dermatitis or for determining whether an individual is at risk of  
 CC developing atopic dermatitis by determining the presence of PRP- or PIP-  
 CC specific antibodies or immune complexes, or by quantifying histamine  
 CC release; and sensitisation remedies for atopic dermatitis containing PRP  
 CC and/or PIP or their peptides as the active ingredient. PRP, PIP and their  
 CC antibodies are useful in diagnosing atopic dermatitis, or for determining  
 CC whether an individual is at risk of developing atopic dermatitis. They  
 CC are also useful in developing sensitisation remedies for the treatment of  
 CC atopic dermatitis. The present sequence represents the specifically  
 CC claimed human salivary acidic proline-rich phosphoprotein (PRP).

XX SQ Sequence 166 AA;

Query Match

Best Local Similarity 100.0%; Score 104; DB 7; Length 166;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGHPRPGRGPPQQ 17

Db 115 GGHPRPGRGPPQQ 131

## RESULT 12

ABP76137

ID ABP76137 standard; protein; 106 AA.

XX AC ABP76137;

XX DT 21-FEB-2003 (first entry)

DE Human GENSET protein SEQ ID 463.  
 XX Cytostatic; antiinflammatory; nootropic; neuroprotective; cardiant;  
 KW gastrointestinal; gene therapy; GENSET; heavy metal toxicity; cancer;  
 KW inflammatory disease; immune disorder; neuromuscular; toxicity;  
 KW central nervous system; cardiovascular; gastrointestinal.  
 XX Homo sapiens.  
 OS  
 XX WO200283898-A1.  
 PN  
 XX 24-OCT-2002.  
 PD  
 XX  
 PF 18-APR-2001; 2001WO-IB000914.  
 XX  
 XX 18-APR-2001; 2001WO-IB000914.  
 PR (GEST ) GENSET.  
 XX  
 PI Bejanin S, Tanaka H, Dumas Milne Edwards J, Jobert S, Giordano J;  
 XX WPI; 2003-075548/07.  
 DR  
 XX New GENSET polynucleotides and polypeptides, useful for treating heavy  
 PT metal toxicity, cancer, inflammatory diseases, immune disorders, and the  
 PT neuromuscular, CNS, cardiovascular or gastrointestinal effects of the  
 PT toxicity.  
 XX  
 PS Claim 14; Page 404-405; 735pp; English.  
 XX  
 CC The present invention relates to novel GENSET polynucleotides (AB236404-  
 CC AB236911) encoding polypeptides (ABP75963-ABP76368). The polynucleotides  
 CC and polypeptides are useful in screening and diagnostic assays for  
 CC abnormal GENSET expression and/or biological activity. They are also  
 CC useful for screening of compounds for treating or preventing GENSET-  
 CC related disorders, such as heavy metal toxicity, cancer, inflammatory  
 CC diseases, immune disorders, and the neuromuscular, central nervous system  
 CC (CNS), cardiovascular or gastrointestinal effects of the toxicity  
 XX  
 SQ Sequence 106 AA;  
 XX  
 Query Match 67.3%; Score 70; DB 6; Length 106;  
 Best Local Similarity 78.6%; Pred. No. 0.37;  
 Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 XX  
 QY 4 PRPRGRPGQPQQ 17  
 Db 84 PPPQGRPGQPQQ 97  
 XX  
 RESULT 13  
 ABP75971  
 ID ABP75971 standard; protein; 106 AA.  
 AC  
 XX ABP75971;  
 XX  
 DT 21-FEB-2003 (first entry)  
 XX  
 DE Human GENSET protein SEQ ID 178.  
 XX  
 KW Cytostatic; antiinflammatory; nootropic; neuroprotective; cardiant;  
 KW gastrointestinal; gene therapy; GENSET; heavy metal toxicity; cancer;  
 KW inflammatory disease; immune disorder; neuromuscular; toxicity;  
 KW central nervous system; cardiovascular; gastrointestinal.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200283898-A1.  
 XX  
 PD 24-OCT-2002.  
 XX  
 PF 18-APR-2001; 2001WO-IB000914.  
 XX

PR 18-APR-2001; 2001WO-IB000914.  
 XX (GEST ) GENSET.  
 XX Bejanin S, Tanaka H, Dumas Milne Edwards J, Jobert S, Giordano J;  
 XX WPI; 2003-075548/07.  
 DR  
 XX New GENSET polynucleotides and polypeptides, useful for treating heavy  
 PT metal toxicity, cancer, inflammatory diseases, immune disorders, and the  
 PT neuromuscular, CNS, cardiovascular or gastrointestinal effects of the  
 PT toxicity.  
 XX  
 PS Claim 14; Page 404-405; 735pp; English.  
 XX  
 CC The present invention relates to novel GENSET polynucleotides (AB236404-  
 CC AB236911) encoding polypeptides (ABP75963-ABP76368). The polynucleotides  
 CC and polypeptides are useful in screening and diagnostic assays for  
 CC abnormal GENSET expression and/or biological activity. They are also  
 CC useful for screening of compounds for treating or preventing GENSET-  
 CC related disorders, such as heavy metal toxicity, cancer, inflammatory  
 CC diseases, immune disorders, and the neuromuscular, central nervous system  
 CC (CNS), cardiovascular or gastrointestinal effects of the toxicity  
 XX  
 SQ Sequence 106 AA;  
 XX  
 Query Match 67.3%; Score 70; DB 6; Length 106;  
 Best Local Similarity 78.6%; Pred. No. 0.37;  
 Matches 11; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
 XX  
 QY 4 PRPRGRPGQPQQ 17  
 Db 84 PPPQGRPGQPQQ 97  
 XX  
 RESULT 14  
 AAW03557  
 ID AAW03557 standard; peptide; 44 AA.  
 AC AAW03557;  
 XX  
 XX 10-DEC-1996 (first entry)  
 DT  
 XX Human proline-rich saliva glycoprotein P-C.  
 DE  
 XX Human; saliva; proline-rich glycoprotein; insulin; secretion; glucagon;  
 KW perfusion pancreas model; Wistar rat; diabetes.  
 KW  
 OS Homo sapiens.  
 XX  
 PN JP08092281-A.  
 XX  
 XX 09-APR-1996.  
 PD  
 XX 27-SEP-1994; 94JP-00257540.  
 PF  
 XX 27-SEP-1994; 94JP-00257540.  
 FR  
 XX (TEIK ) TEIKOKU HORMONE MFG CO LTD.  
 PA  
 XX WPI; 1996-236094/24.  
 DR  
 XX New octadecapeptide for treating diabetes - acts as insulin secretion  
 PT promoter and glucagon secretion inhibitor.  
 PT  
 XX Disclosure; Page 2; 7pp; Japanese.  
 PS  
 XX This is the sequence of the human proline-rich saliva glycoprotein P-C  
 CC which was used to synthesize an octadecapeptide (AAW03556) which has an  
 CC insulin secretion promoting activity and a glucagon secretion inhibiting  
 CC activity. The peptide was generated by usual peptide synthesis inhibiting  
 CC opt. contains Boc protective gps. at the N-terminal His and on the  
 CC residue during synthesis. These are removed by cleavage with HCl for 10

CC mins on ice. The insulin secretion promoting activity and glucagon  
 CC secretion inhibiting activity were measured in a perfusion pancreas model  
 CC in a male Wistar rat. Insulin secretion was 1.78 times than normal and  
 CC glucagon secretion was 0.77 times normal levels. The peptide can thus be  
 CC used to treat diabetes

XX  
 SQ Sequence 44 AA;

Query Match 66.3%; Score 69; DB 2; Length 44;  
 Best Local Similarity 61.9%; Pred. No. 0.22;  
 Matches 13; Conservative 1; Mismatches 3; Indels 4; Gaps 1;

OY 1 GGH----PRPRGRPGPPQQ 17  
 ||| ||| :|||  
 Db 10 GGHQGGPPPPPPGKPGPPQ 30

## RESULT 15

AAP93320  
 ID AAP93320 standard; protein; 45 AA.

XX  
 AC AAP93320;

XX  
 DT 19-JUN-1990 (first entry)

XX  
 DE P-C gene.

XX  
 KW P-C gene; fused protein; silica gel absorbent.

XX  
 OS Unidentified.

XX  
 PN JP01051097-A.

XX  
 PD 27-FEB-1989.

XX  
 PF 21-AUG-1987; 87JP-00208616.

XX  
 PR 11-DEC-1979; 79JP-00160528.

XX  
 PA (TORA ) TORAY IND INC.

XX  
 DR WPI; 1989-104005/14.

XX  
 DR N-PSDB; AAN92373.

XX  
 PT Purificn. of fused protein(s) - comprises contacting fused protein with  
 PT silica gel absorbent and eluting.

XX  
 PS Fig 1; page 6; 8pp; Japanese.

XX  
 CC The DNA encoding it is pred. as 14 oligomers. The 5' terminal of 12 of  
 CC the oligomers is phosphorylated, complementary 2 or 3 oligomers are  
 CC annealed, and the resulting 6 ds DNAs are combined to form the P-C gene

XX  
 SQ Sequence 45 AA;

Query Match 66.3%; Score 69; DB 1; Length 45;  
 Best Local Similarity 61.9%; Pred. No. 0.23;  
 Matches 13; Conservative 1; Mismatches 3; Indels 4; Gaps 1;

OY 1 GGH----PRPRGRPGPPQQ 17  
 ||| ||| :|||  
 Db 11 GGHQGGPPPPPPGKPGPPQ 31

Search completed: April 6, 2004, 16:06:42  
 Job time : 95.1682 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 16:14:50 ; Search time 66.729 Seconds  
(without alignments)  
66.909 Million cell updates/sec

Title: US-10-009-709-13

Perfect score: 104

Sequence: 1 GGHPRPRGRPQGPPQQ 17

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1071772 seqs, 26263353 residues

Total number of hits satisfying chosen parameters: 1071772

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	104	100.0	132	9	US-09-864-761-43644
2	104	100.0	132	14	US-10-157-031-80
3	58	55.8	19652	15	US-10-084-846A-7
4	56	53.8	191	15	US-10-374-780A-1418
5	56	53.8	310	14	US-10-156-761-12127
6	56	53.8	398	12	US-10-425-114-61067
7	56	53.8	19723	15	US-10-084-846A-5
8	55.5	53.4	953	9	US-09-888-615-66
9	55.5	53.4	933	16	US-10-311-035-21
10	55	52.9	123	12	US-10-424-599-259584
11	55	52.9	149	12	US-10-424-599-171613
12	55	52.9	240	12	US-10-425-114-68908
13	55	52.9	299	12	US-10-425-114-40613
14	55	52.9	539	12	US-10-425-114-39176
15	54.5	52.4	52	12	US-10-424-599-277859

16	54	51.9	128	12	US-10-424-599-177399
17	54	51.9	155	12	US-10-424-599-171668
18	54	51.9	162	12	US-10-424-599-219187
19	54	51.9	245	8	US-08-424-550B-38
20	54	51.9	269	12	US-10-424-599-206111
21	54	51.9	641	10	US-09-959-987-2
22	54	51.9	647	9	US-09-925-299-1002
23	54	51.9	647	10	US-09-925-299-1002
24	53	51.0	107	11	US-09-833-245-280
25	53	51.0	131	12	US-10-424-599-185075
26	53	51.0	551	11	US-09-764-875-892
27	53	51.0	707	10	US-09-919-039-278
28	53	51.0	2017	12	US-10-114-270-86
29	53	51.0	19662	15	US-10-084-846A-6
30	52.5	50.5	165	12	US-10-425-114-70160
31	52	50.0	64	10	US-09-933-767-1109
32	52	50.0	64	14	US-10-023-282-1109
33	52	50.0	72	12	US-10-424-599-148528
34	52	50.0	94	12	US-10-424-599-189787
35	52	50.0	94	12	US-10-424-599-285441
36	52	50.0	95	12	US-10-424-599-239587
37	52	50.0	230	12	US-10-424-599-233437
38	52	50.0	451	14	US-10-127-940-2
39	52	50.0	451	14	US-10-441-757-2
40	52	50.0	452	10	US-09-443-745-16
41	52	50.0	1008	12	US-10-222-772-62
42	52	50.0	1028	15	US-10-369-493-3464
43	52	50.0	3122	14	US-10-200-562-201
44	52	50.0	3122	14	US-10-237-551-201
45	52	50.0	3122	14	US-10-237-551-250

#### ALIGNMENTS

#### RESULT 1

US-09-864-761-43644  
; Sequence 43644, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
; FILE REFERENCE: Aomics-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: US 09/632,366  
; PRIOR FILING DATE: 2000-08-03  
; PRIOR APPLICATION NUMBER: GB 24263.5  
; PRIOR FILING DATE: 2000-10-04  
; PRIOR APPLICATION NUMBER: US 60/236,359  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: PCT/US01/006666  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006667  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006664  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006669  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006665  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006668  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: PCT/US01/006663  
; PRIOR FILING DATE: 2001-01-30



PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 43644  
LENGTH: 132  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC006518.17  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.95  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.77  
OTHER INFORMATION: EST HUMAN HIT: BF08785.1, EVALUE 1.00e-01  
OTHER INFORMATION: SWISSPROT HIT: P02810, EVALUE 7.40e-02  
US-09-864-761-43644

Query Match 100.0%; Score 104; DB 9; Length 132;  
Best Local Similarity 100.0%; Pred. No. 0.00033;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGHPRRPRGRPQGPPQQ 17  
Db 81 GGHPRRPRGRPQGPPQQ 97

RESULT 2  
US-10-157-031-80  
Sequence 80, Application US/10157031  
Publication No. US2003010890A1  
GENERAL INFORMATION:  
APPLICANT: Baranova, A. V.  
APPLICANT: Yankovsky, N. K.  
APPLICANT: Kozlov, A. P.  
APPLICANT: Lobashev, A. V.  
APPLICANT: Krukovskaya, L. L.  
TITLE OF INVENTION: In silico screening for phenotype-associated expressed sequences  
FILE REFERENCE: 2760-103  
CURRENT APPLICATION NUMBER: US/10/157,031  
CURRENT FILING DATE: 2002-05-30  
NUMBER OF SEQ ID NOS: 415  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 80  
LENGTH: 166  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-157-031-80

Query Match 100.0%; Score 104; DB 14; Length 166;  
Best Local Similarity 100.0%; Pred. No. 0.0004;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGHPRRPRGRPQGPPQQ 17  
Db 115 GGHPRRPRGRPQGPPQQ 131

RESULT 3  
US-10-084-846A-7  
Sequence 7, Application US/10084846A

Publication No. US20040006026A1  
GENERAL INFORMATION:  
APPLICANT: WEITNAUER, GABRIELE  
APPLICANT: MUHLENWEG, AGNES  
APPLICANT: TREFFZER, AXEL  
APPLICANT: BECHTHOLD, ANDREAS  
TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
FILE REFERENCE: 1974-005  
CURRENT APPLICATION NUMBER: US/10/084,846A  
CURRENT FILING DATE: 2003-02-25  
PRIOR APPLICATION NUMBER: PCT/EP01/09815  
PRIOR FILING DATE: 2001-08-24  
PRIOR APPLICATION NUMBER: DE 101 09 166.4  
PRIOR FILING DATE: 2001-02-25  
NUMBER OF SEQ ID NOS: 120  
SOFTWARE: PatentIn Ver. 3.2  
SEQ ID NO 7  
LENGTH: 19652  
TYPE: PRT  
ORGANISM: Streptomyces viridochromogenes  
FEATURE:  
OTHER INFORMATION: Protein 2: amino acid sequence encoded by coding strand 2.  
OTHER INFORMATION: Start codon: gat, Start position: nucleotide 2.  
US-10-084-846A-7

Query Match 55.8%; Score 58; DB 15; Length 19652;  
Best Local Similarity 76.9%; Pred. No. 2.2e-03;  
Matches 10; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 2 GHPRRPRGRPQQ 14  
Db 2258 GTPRRPRGRGRGP 2270

RESULT 4  
US-10-374-780A-1418  
Sequence 1418, Application US/10374780A  
Publication No. US20040019927A1  
GENERAL INFORMATION:  
APPLICANT: Sherman, Bradley K  
APPLICANT: Riechmann, Jose Luis  
APPLICANT: Jiang, Cai-Zhong  
APPLICANT: Heard, Jacqueline E  
APPLICANT: Haake, Volker  
APPLICANT: Creelman, Robert A  
APPLICANT: Ratcliffe, Oliver  
APPLICANT: Adam, Luc J  
APPLICANT: Reuber, T. Lynne  
APPLICANT: Keddie, James  
APPLICANT: Brown, Pierre E  
APPLICANT: Pilgrim, Marsha L  
APPLICANT: Dubell III, Arnold T  
APPLICANT: Pineda, Omaira  
APPLICANT: Yu, Guo-Liang  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS  
FILE REFERENCE: MBI-0047 CIP  
CURRENT APPLICATION NUMBER: US/10/374,780A  
CURRENT FILING DATE: 2003-02-25  
PRIOR APPLICATION NUMBER: 09/837,944  
PRIOR FILING DATE: 2001-04-18  
PRIOR APPLICATION NUMBER: 60/310,847  
PRIOR FILING DATE: 2001-08-09  
PRIOR APPLICATION NUMBER: 09/934,455  
PRIOR FILING DATE: 2001-08-22  
PRIOR APPLICATION NUMBER: 60/336,049  
PRIOR FILING DATE: 2001-11-19  
PRIOR APPLICATION NUMBER: 60/338,692  
PRIOR FILING DATE: 2001-12-11  
PRIOR APPLICATION NUMBER: 10/171,468  
PRIOR FILING DATE: 2002-06-14  
PRIOR APPLICATION NUMBER: 10/225,066  
PRIOR FILING DATE: 2002-08-09  
PRIOR APPLICATION NUMBER: 10/225,067

; PRIOR FILING DATE: 2002-08-09  
; PRIOR APPLICATION NUMBER: 10/225,068  
; PRIOR FILING DATE: 2002-08-09  
; NUMBER OF SEQ ID NOS: 2906  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 1418  
; LENGTH: 191  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Orthologous to GI073  
US-10-374-780A-1418

Query Match 53.8%; Score 56; DB 15; Length 191;  
Best Local Similarity 57.1%; Pred. No. 68;  
Matches 12; Conservative 0; Mismatches 3; Indels 5; Gaps 1;

QY 1 GGHPRPPGRPQG-----PP 15  
Db 50 GGGRRRPRGRPPGSKNPKPP 70

## RESULT 5

US-10-156-761-12127  
; Sequence 12127, Application US/10156761  
; Publication No. US20030119018A1  
; GENERAL INFORMATION:  
; APPLICANT: OMURA, SATOSHI  
; APPLICANT: IKEDA, HARUO  
; APPLICANT: ISHIKAWA, JUN  
; APPLICANT: HORIKAWA, HIROSHI  
; APPLICANT: SHIBA, TADAYOSHI  
; APPLICANT: SAKAKI, YOSHIYUKI  
; APPLICANT: HATTORI, MASAHIRA  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-262  
; CURRENT APPLICATION NUMBER: US/10/156,761  
; CURRENT FILING DATE: 2002-05-29  
; PRIOR APPLICATION NUMBER: JP 2001-204089  
; PRIOR FILING DATE: 2001-05-30  
; PRIOR APPLICATION NUMBER: JP 2001-272697  
; PRIOR FILING DATE: 2001-08-02  
; NUMBER OF SEQ ID NOS: 15109  
; SEQ ID NO 12127  
; LENGTH: 310  
; TYPE: PRT  
; ORGANISM: Streptomyces avermitilis  
US-10-156-761-12127

Query Match 53.8%; Score 56; DB 14; Length 310;  
Best Local Similarity 66.7%; Pred. No. 1e+02;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GGHPRPPGRPQGPP 15  
Db 296 GGIPAPPLGAPQGP 310

## RESULT 6

US-10-425-114-61067  
; Sequence 61067, Application US/10425114  
; Publication No. US20040034888A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiu, Jingdong  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Kovacic, David K.  
; APPLICANT: Screen, Steven E  
; APPLICANT: Tabaska, Jack E  
; APPLICANT: Cao, Yongwei  
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with  
; FILE REFERENCE: 38-21/53313/B  
; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 73128  
; SEQ ID NO 61067  
; LENGTH: 398  
; TYPE: PRT  
; ORGANISM: Zea mays  
; FEATURE:  
; OTHER INFORMATION: Clone ID: LIB3279-185-B11\_FLI.pep  
US-10-425-114-61067

Query Match 53.8%; Score 56; DB 12; Length 398;  
Best Local Similarity 66.7%; Pred. No. 1.3e+02;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GGHPRPPGRPQGPP 15  
Db 196 GHEPRAPLGRFPQDP 210

## RESULT 7

US-10-084-846A-5  
; Sequence 5, Application US/10084846A  
; Publication No. US20040008028A1  
; GENERAL INFORMATION:  
; APPLICANT: WEITNAUER, GABRIELE  
; APPLICANT: MUHLENWEG, AGNES  
; APPLICANT: TREPZER, AXEL  
; APPLICANT: BECHTHOLD, ANDREAS  
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES  
; FILE REFERENCE: 1974-005  
; CURRENT APPLICATION NUMBER: US/10/084,846A  
; CURRENT FILING DATE: 2003-02-25  
; PRIOR APPLICATION NUMBER: PCT/EP01/09815  
; PRIOR FILING DATE: 2001-08-24  
; PRIOR APPLICATION NUMBER: DE 101 09 166.4  
; PRIOR FILING DATE: 2001-02-25  
; NUMBER OF SEQ ID NOS: 120  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 5  
; LENGTH: 19723  
; TYPE: PRT  
; ORGANISM: Streptomyces viridochromogenes  
; FEATURE:  
; OTHER INFORMATION: Protein 3: amino acid sequence encoded by coding strand 1.  
; OTHER INFORMATION: Start codon: atc, Start position: nucleotide 3.  
US-10-084-846A-5

Query Match 53.8%; Score 56; DB 15; Length 19723;  
Best Local Similarity 60.0%; Pred. No. 3.6e+03;  
Matches 12; Conservative 1; Mismatches 3; Indels 4; Gaps 1;

QY 1 GGH---PRPPRGRPQGPPQ 16  
Db 10177 GGHRTSGRRRPRGRPQ 10196

## RESULT 8

US-09-888-615-66  
; Sequence 66, Application US/09888615  
; Patent No. US20020064856A1  
; GENERAL INFORMATION:  
; APPLICANT: FLOWMAN, GREGORY  
; APPLICANT: WHYTE, DAVID  
; APPLICANT: CAENEPEEL, SEAN  
; APPLICANT: CHARYDCZAK, GLEN  
; APPLICANT: MANNING, GERARD  
; APPLICANT: SUDARSANAM, SUCHA  
; TITLE OF INVENTION: NOVEL PROTEASES  
; FILE REFERENCE: 038602/1214  
; CURRENT APPLICATION NUMBER: US/09/888,615  
; CURRENT FILING DATE: 2001-06-26  
; PRIOR APPLICATION NUMBER: 60/214,047  
; PRIOR FILING DATE: 2000-06-26

```
; NUMBER OF SEQ ID NOS: 150
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 66
; LENGTH: 953
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-888-615-66

Query Match      53.4%; Score 55.5; DB 9; Length 953;
Best Local Similarity 57.9%; Pred. No. 3e+02;
Matches 11; Conservative 1; Mismatches 4; Indels 3; Gaps 1;

QY 1 GGHRP---PPRGPPQPPQ 16
      |||||
Db 217 GGHRSAWPPRHAQWPPE 235

RESULT 9
US-10-311-035-21
; Sequence 21, Application US/10311035
; Publication No. US20040023243A1
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: YUE, Henry
; APPLICANT: ELLIOTT, Vicki
; APPLICANT: GANDHI, Ameena R.
; APPLICANT: LAL, Preeti
; APPLICANT: AU-YOUNG, Janice
; APPLICANT: TRIBOULEY, Catherine M.
; APPLICANT: DELEGEANE, Angelo M.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: NGUYEN, Dannie B.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: HAPALIA, April
; APPLICANT: KHAN, Farrah A.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: YAO, Monique G.
; APPLICANT: LU, Dyung Aina M.
; APPLICANT: ARVIZU, Chandra S.
; APPLICANT: TANG, Y. Tom
; APPLICANT: WALSH, Roderick T.
; APPLICANT: AZIMZAI, Yalda
; APPLICANT: LU, Yan
; APPLICANT: RAMKUMAR, Jayalaximi
; APPLICANT: XU, Yuming
; APPLICANT: REDDY, Roopa
; APPLICANT: DAS, Depotriya
; APPLICANT: KEARNEY, Liam
; APPLICANT: KALLICK, Deborah A.
; TITLE OF INVENTION: Proteases
; FILE REFERENCE: FI-0123 PCT
; CURRENT APPLICATION NUMBER: US/10/311,035
; CURRENT FILING DATE: 2002-12-10
; PRIOR APPLICATION NUMBER: 60/212,336; 60/213,955; 60/215,396; 60/216,821; 60/218,946
; PRIOR FILING DATE: 2000-06-16; 2000-06-22; 2000-06-29; 2000-07-07; 2000-07-14
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PERL Program
; SEQ ID NO 21
; LENGTH: 953
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20040023243A1 6817347CD1
US-10-311-035-21

Query Match      53.4%; Score 55.5; DB 16; Length 953;
Best Local Similarity 57.9%; Pred. No. 3e+02;
Matches 11; Conservative 1; Mismatches 4; Indels 3; Gaps 1;

QY 1 GGHRP---PPRGPPQPPQ 16
      |||||
Db 217 GGHRSAWPPRHAQWPPE 235

US-10-424-599-259584
; Sequence 259584, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 259584
; LENGTH: 123
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: unsure
; LOCATION: (1)..(123)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_7642C.1.pep
US-10-424-599-259584

Query Match      52.9%; Score 55; DB 12; Length 123;
Best Local Similarity 64.7%; Pred. No. 59;
Matches 11; Conservative 1; Mismatches 3; Indels 2; Gaps 1;

QY 1 GGHRP--PRGRPGPP 15
      |||||
Db 82 GGPRPPLQPSGRPKPP 98

RESULT 11
US-10-424-599-171613
; Sequence 171613, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 171613
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Glycine max
; NAME/KEY: misc feature
; OTHER INFORMATION: Clone ID: PAT_MRT3847_125982C.1.pep
US-10-424-599-171613

Query Match      52.9%; Score 55; DB 12; Length 149;
Best Local Similarity 58.8%; Pred. No. 70;
Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GGHRP--PRGRPGPPQ 17
      |||||
Db 48 GKRDTEKGRPGGPPQK 64

RESULT 12
US-10-425-114-68908
; Sequence 68908, Application US/10425114
```

```
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 68908
; LENGTH: 240
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLMO17007B11_FLI.pep
US-10-425-114-68908

Query Match          52.9%; Score 55; DB 12; Length 240;
Best Local Similarity 65.0%; Pred. NO. 1.1e+02;
Matches 13; Conservative 1; Mismatches 2; Indels 4; Gaps 2;

QY      2 GHRPPP---RGRP-QGPPQQ 17
Db      155 GHRPPPLDRQRPQQGPPRQ 174

RESULT 13
US-10-425-114-40613
; Sequence 40613, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 40613
; LENGTH: 299
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB189-030-B4_FLI.pep
US-10-425-114-40613

Query Match          52.9%; Score 55; DB 12; Length 299;
Best Local Similarity 65.0%; Pred. NO. 1.3e+02;
Matches 13; Conservative 1; Mismatches 2; Indels 4; Gaps 2;

QY      2 GHRPPP---RGRP-QGPPQQ 17
Db      214 GHRPPPLDRQRPQQGPPRQ 233

RESULT 14
US-10-425-114-39176
; Sequence 39176, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 39176
; LENGTH: 539
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3245-443-A8_FLI.pep
US-10-425-114-39176

Query Match          52.9%; Score 55; DB 12; Length 539;
Best Local Similarity 65.0%; Pred. NO. 2.1e+02;
Matches 13; Conservative 1; Mismatches 2; Indels 4; Gaps 2;

QY      2 GHRPPP---RGRP-QGPPQQ 17
Db      454 GHRPPPLDRQRPQQGPPRQ 473

RESULT 15
US-10-424-599-277859
; Sequence 277859, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 277859
; LENGTH: 52
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_9292C.1.pep
US-10-424-599-277859

Query Match          52.4%; Score 54.5; DB 12; Length 52;
Best Local Similarity 73.3%; Pred. NO. 32;
Matches 11; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY      2 GHRP-PPRGRPQQGP 15
Db      2 GPRGPPRGPPGPP 16

Search completed: April 6, 2004, 17:06:10
Job time : 67.729 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 6, 2004, 15:56:34 ; Search time 24.9439 Seconds  
(without alignments)  
35.185 Million cell updates/sec

Title: US-10-009-709-13

Perfect score: 104

Sequence: 1 GGHPRPRGRPGPPQ 17

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2.6/prodata/2/1aa/5A-COMB.pap:\*

2: /cgn2.6/prodata/2/1aa/5B-COMB.pap:\*

3: /cgn2.6/prodata/2/1aa/6A-COMB.pap:\*

4: /cgn2.6/prodata/2/1aa/6B-COMB.pap:\*

5: /cgn2.6/prodata/2/1aa/PCITUS-COMB.pap:\*

6: /cgn2.6/prodata/2/1aa/backfiles1.pap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	63	60.6	377	4	US-09-252-991A-29332
2	62	59.6	82	2	US-08-925-237-4
3	62	59.6	124	2	US-08-925-237-2
4	56	53.8	326	4	US-09-252-991A-30870
5	56	53.8	559	4	US-09-252-991A-24480
6	56	53.8	1213	3	US-09-413-814-79
7	55	52.9	223	4	US-09-252-991A-17072
8	55	52.9	1706	4	US-09-252-991A-31760
9	54	51.9	27	3	US-08-658-136-29
10	54	51.9	245	4	US-08-469-260A-38
11	54	51.9	245	4	US-08-488-446-38
12	54	51.9	245	4	US-08-467-344A-38
13	54	51.9	391	4	US-09-252-991A-20316
14	54	51.9	422	4	US-09-252-991A-27513
15	54	51.9	453	4	US-09-252-991A-27041
16	54	51.9	1614	4	US-09-052-469-2
17	54	51.9	1614	4	US-08-422-582-2
18	54	51.9	1614	4	US-09-052-262-2
19	54	51.9	4339	4	US-09-052-469-6
20	54	51.9	4339	4	US-08-422-582-6
21	54	51.9	4339	4	US-09-052-262-6
22	53.5	51.4	97	4	US-09-489-039A-8140
23	53.5	51.4	152	4	US-09-252-991A-18309
24	53.5	51.4	342	4	US-09-252-991A-21589
25	53	51.0	142	4	US-09-252-991A-21936
26	53	51.0	276	4	US-09-252-991A-23623
27	53	51.0	515	4	US-09-252-991A-28127

Sequence 26622, A  
Sequence 25595, A  
Sequence 1109, Ap  
Sequence 29505, A  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 16, Appli  
Sequence 16, Appli  
Sequence 2621, A  
Sequence 26802, A  
Sequence 17018, A  
Sequence 29074, A  
Sequence 4, Appli  
Sequence 4, Appli  
Sequence 26318, A  
Sequence 21709, A

US-09-252-991A-26622  
US-09-252-991A-25595  
US-09-205-258-1109  
US-09-252-991A-29505  
US-08-570-157-2  
US-09-076-510-2  
US-09-004-349-2  
US-07-937-609-16  
US-08-029-170-16  
US-09-252-991A-32621  
US-09-252-991A-26802  
US-09-252-991A-17018  
US-09-252-991A-29074  
US-09-026-587-4  
US-09-227-420-4  
US-08-387-811-4  
US-09-252-991A-26318  
US-09-252-991A-21709

#### ALIGNMENTS

##### RESULT 1

US-09-252-991A-29332  
; Sequence 29332, Application US/09252991A  
; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 29332

; LENGTH: 377

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-29332

Query Match 60.6%; Score 63; DB 4; Length 377;  
Best Local Similarity 73.3%; Pred. No. 2.1;  
Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 GHHPRPRGRPGPPQ 16

DB 91 GCHPRPRGRPGPPQ 105

##### RESULT 2

US-08-925-237-4

; Sequence 4, Application US/08925237

; Patent No. 5981720

; GENERAL INFORMATION:

; APPLICANT: Azen, Edwin A.

; APPLICANT: Pan, David

; TITLE OF INVENTION: Human Salivary Proteins And Fragments

; TITLE OF INVENTION: Thereof Having Alpha-Glucosidase Inhibitory Activity

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Kent Bara

; STREET: 100 East Wisconsin Avenue

; CITY: Milwaukee

; STATE: WI USA

; COUNTRY: USA

; ZIP: 53202-4108

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

Query Match 59.6%; Score 62; DB 2; Length 124;

US-09-413-814-79  
; Sequence 79, Application US/09413814

Patent No. 6225064  
GENERAL INFORMATION:  
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH  
APPLICANT: Bristol-Myers Squibb, Co.  
APPLICANT: Beyer, Stefan  
APPLICANT: Bloecker, Helmut  
APPLICANT: Brandt, Petra  
APPLICANT: Cino, Paul M  
APPLICANT: Dougherty, Brian A  
APPLICANT: Goldberg, Steven L  
APPLICANT: Hofle, Gerhard  
APPLICANT: Mueller, Joachim  
APPLICANT: Reichenbach, Hans  
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or heteropolyketide compounds  
FILE REFERENCE: PCT/US 99/23535  
CURRENT APPLICATION NUMBER: US/09/413,814  
CURRENT FILING DATE: 1999-10-07  
EARLIER APPLICATION NUMBER: DE 198 46 493.2  
EARLIER FILING DATE: 1998-10-09  
NUMBER OF SEQ ID NOS: 107  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 79  
LENGTH: 1213  
TYPE: PRT  
ORGANISM: Sorangium cellulosum  
US-09-413-814-79

Query Match 53.8%; Score 56; DB 3; Length 1213;  
Best Local Similarity 75.0%; Pred. No. 39;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 PRPPRGRPGQP 15  
Db 66 PSPRGAQAPP 77

RESULT 7  
US-09-252-991A-17072  
Sequence 17072, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 17072  
LENGTH: 223  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17072

Query Match 52.9%; Score 55; DB 4; Length 223;  
Best Local Similarity 90.0%; Pred. No. 11;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GHRPPRGRP 11  
Db 138 GHERPPGRP 147

RESULT 8  
US-09-252-991A-31760  
Sequence 31760, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
PRIOR FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 17072  
LENGTH: 223  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-31760

Query Match 52.9%; Score 55; DB 4; Length 1706;  
Best Local Similarity 66.7%; Pred. No. 69;  
Matches 10; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 GHRPPRGRPGQP 16  
Db 1545 GRPPPPGRAPGRQ 1559

RESULT 9  
US-08-658-136-29  
Sequence 29, Application US/08658136  
Patent No. 6071717  
GENERAL INFORMATION:  
APPLICANT: KLINGER, KATHERINE W  
APPLICANT: LANDES, GREGORY M  
APPLICANT: BURN, TIMOTHY C  
APPLICANT: CONNORS, TIMOTHY D  
APPLICANT: DACKOWSKI, WILLIAM  
APPLICANT: GERMING, GREGORY  
APPLICANT: QIAN, FENG  
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENZYME CORPORATION  
STREET: ONE MOUNTAIN ROAD  
CITY: FRAMINGHAM  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 01701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/658,136  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: LASSEN, ELIZABETH  
REGISTRATION NUMBER: 31,845  
REFERENCE/DOCKET NUMBER: GEN4-17.8  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 508-872-8400  
TELEFAX: 508-872-5415  
INFORMATION FOR SEQ ID NO: 29:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 27 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-658-136-29

Query Match 51.9%; Score 54; DB 3; Length 27;  
Best Local Similarity 62.5%; Pred. No. 2.2;





TITLE OF INVENTION: NON-A, NON-B, NON-C, NON-D, NON-E HEPATITIS  
REAGENTS AND METHODS FOR THEIR USE

NUMBER OF SEQUENCES: 716

CORRESPONDENCE ADDRESS:

ADDRESSEE: ABBOTT LABORATORIES D377/AP6D

STREET: 100 ABBOTT PARK ROAD

CITY: ABBOTT PARK

STATE: IL

COUNTRY: USA

ZIP: 60064-3500

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/467,344A

FILING DATE: 07-Jun-1995

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/424,550

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: POROMBSKI, PRISCILLA E.

REGISTRATION NUMBER: 33,207

REFERENCE/DOCKET NUMBER: 5527.PC.01

TELEPHONE: 708-937-6365

TELEFAX: 708-938-2623

INFORMATION FOR SEQ ID NO: 38:

SEQUENCE CHARACTERISTICS:

LENGTH: 245 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 38:

US-08-467-344A-38

Query Match 51.9%; Score 54; DB 4; Length 245;

Best Local Similarity 69.2%; Pred. No. 16;

Matches 9; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 HRPGRGRPGGPP 15

Db 122 HRPGRGRPGGPP 134

RESULT 13

US-09-252-991A-20316

Sequence 20316, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 20316

LENGTH: 391

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-20316

Query Match 51.9%; Score 54; DB 4; Length 391;

Best Local Similarity 45.8%; Pred. No. 24;

Matches 11; Conservative 2; Mismatches 3; Indels 8; Gaps 1;

QY 2 GHP-----RPPGRPGGPPQ 17

Db 91 GHPGVVADTLRQHRGRPGGPP 114

RESULT 14

US-09-252-991A-27513

Sequence 27513, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 27513

LENGTH: 422

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-27513

Query Match 51.9%; Score 54; DB 4; Length 422;

Best Local Similarity 64.3%; Pred. No. 26;

Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 3 HRPGRGRPGGPP 16

Db 376 HPDPRGQHRRPQ 389

RESULT 15

US-09-252-991A-27041

Sequence 27041, Application US/09252991A

Patent No. 6551795

GENERAL INFORMATION:

APPLICANT: Marc J. Rubenfield et al.

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

FILE REFERENCE: 107196.136

CURRENT APPLICATION NUMBER: US/09/252,991A

PRIOR FILING DATE: 1999-02-18

PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18

PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27

NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 27041

LENGTH: 453

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-27041

Query Match 51.9%; Score 54; DB 4; Length 453;

Best Local Similarity 60.0%; Pred. No. 27;

Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 GHPGRGRPGGPP 16

Db 270 GHPGRGRPGGPP 284

Search completed: April 6, 2004, 16:19:42

Job time : 24.9439 secs